Examination of insurer fees by means of life expectancy

(Nederlandse titel: Onderzoek naar verzekeraarstarieven aan de hand van levensverwachting)

Verslag ten behoeve van het Delft Institute of Applied Mathematics als onderdeel ter verkrijging van de graad van

BACHELOR OF SCIENCE in TECHNISCHE WISKUNDE

door

Joost Verkade

Delft, Nederland
August 20, 2014

Copyright © 2014 door Joost Verkade. Alle rechten voorbehouden.
“Examination of insurer fees
by means of life expectancy”

(Nederlandse titel: “Onderzoek naar verzekeraarstarieven
aan de hand van levensverwachting”)

Joost Verkade

Technische Universiteit Delft

Begeleider
D. Kurowicka

Overige commissieleden
A.W. Heemink  E.M. van Elderen

August 20, 2014  Delft
## Contents

1 Introduction 7

2 Pension 9
   2.1 Calculations of the insurer 9
   2.2 Discount rate 10

3 Examination of the data 13
   3.1 Mortality rates 13
   3.2 Maximum age 14
   3.3 Uncertainty in the data 16

4 Life expectancy 19
   4.1 Developments in life expectancy 19

5 AG model 21
   5.1 Approach of the model 21
      5.1.1 Definitions 21
   5.2 Smoothing 21
      5.2.1 Van Broekhoven Algorithm 22
      5.2.2 Exponential smoothing 24
   5.3 Mortality rates for high ages 26
   5.4 The forecasting model 27
      5.4.1 Long and short term trends 27
      5.4.2 Forecasting the mortality rates 30
      5.4.3 Weighted term forecast 32

6 PLT Workgroup 35
   6.1 Reduction factors 35
   6.2 Forecasted mortality rates 36

7 Performance analysis 37

8 Comparison and construction of insurer fees 39
   8.1 Insurer fees 39
   8.2 The use of forecasted fees 40
      8.2.1 Fee comparison 40
   8.3 Using alternative approaches 42
      8.3.1 Retirement tables 42
      8.3.2 Diagonal table 44
   8.4 The benefits of alternative methods 44
1 — Introduction

In the Netherlands, the pension regulation consists of three pillars. The first pillar consists of the basic state pension called AOW (algemene ouderdomswet), which is meant to provide a financial foundation for every person. At the moment, the yearly AOW contribution for a retired person is somewhere between 10,000 and 15,000, depending on whether or not you have children or a partner with income. The second pillar is meant for every employed person as an addition to the AOW pension. Mostly, the employer has an agreement with an insurer, to which he pays a yearly premium for every employee. The insurer then pledges the employee a pension at retirement, amount of which is based on the yearly premium. This premium is partially paid by the employee’s salary and partially by the employer. So the employee actually sets aside a part of salary to save for his pension. Every employer has its own regulation for this secondary term of employment. It is however legally obligatory. The third pillar consists of every other form of building up pension. For example, depositing money into a bank or taking additional insurances.

In this report we will only examine the second pillar. For convenience, we shall simply call this second pillar: pension.

The insurers costs of providing a pension depends on how much and long a person will receive a pension as well as how long this person saves for his/her pension.

The goal of this research is to investigate how the main insurers in the Netherlands determine their costs of providing a pension. We will examine two main algorithms that are used to determine insurers’ premium, discuss assumptions used in these algorithms and investigate whether these assumptions are reasonable. The report is organized as follows:

First, we discuss the general approach how insurers calculate their premiums which are based on the estimated future expenses of the insurer. The future expenses of a person are determined on basis of how long this person is going to live after his retirement and the interest rates. To calculate how much a person has to pay now to receive a certain amount of pension in the future, the calculated expenses are discounted. Since the money put aside now could be invested, it is worth now less than it will be worth in the future. Discounting requires the choice of the interest rate. This discount rate is based on the expected return of risk-free investments such as government bonds. Because the discount rate have a huge impact on the premium, we will discuss it in this report in Section 2.

Every person is different, so the insurer has to estimate the expected life length of a person after retirement. In section 4 the life expectancy is calculated using mortality rates, which provide the probability of death at a certain age. The data about the historical mortality rates of Dutch population is available. However, to calculate future expenses the insurers need the future mortality rates rather than the historical ones. Hence the future mortality rates have to be forecasted.

We will examine two algorithms that allow to forecast mortality rates and provide the life expectancy curves that can be obtain with these forecasted mortality rates. In Section 5 the Royal Actuarial Association (AG) is presented. Section 6 contains explanation of the PLT model (pensioenen lijfrentetafel workgroup). Both models use a similar approach based on the
smooth historical data, which is extrapolated to determine a forecast of 50 years. The models
do not contain uncertainty of the forecast.

An alternative model is proposed by Central Bureau of Statistics (CBS). The CBS model is
different from the other two. It is a lot more complex as it includes causes of death and takes
medical developments into account. Besides that, it does give an indication of the uncertainty in
the forecast, which is a nice advantage. However, due to its complexity, it will not be investigated
in this report.

Section 8 contains comparison of fees that the insurers use to cover their future expenses.
Insurer fees are essentially the premiums for buying €1 of pension of an insurer. Insurers
determine these fees based on their own methods. Some just use the forecast of one of the
earlier mentioned models, while others do their own calculations.

Since the fees would be different for each employee in general the simplified calculations are
performed by insurers. Instead of looking at every employee separately, the insurers calculate
fees for the employee with average age. We will see how these simplified calculations compare
to previously discussed ones.

Finally we conclude the report and provide the recommendation. The detailed presentation
of algorithms discussed in this report are shown in Appendices. Moreover R-code used to make
figures and obtained results included in this report is included in the end.
2 — Pension

In this section the theory of pension will be discussed. It is important to understand how the insurer determines the amount of money needed to make yearly payments for the rest of a person’s life. Since an insurer does not want to lose money, their prices are calculated accurately. These prices are simply based on the expected expenses they have to make.

Employers pay insurers yearly premiums for every employee. In return, the insurer guarantees a yearly contribution for the employee. The employee receives this contribution every year, starting from his retirement age up until his death. The premiums are different for every employee and are depending on the salary, gender and age of the employee.

2.1 Calculations of the insurer

The insurer has to make sure there is enough money to pay the future yearly contributions to an employee. However, since the insurer never knows how old a person will get, he will never know the exact amount of money that is needed. To make an estimation of this amount, the probability of each payout is calculated.

The estimated amount of money is saved up by the employee during his career by paying premiums. Since the insurer invests the money of the employee, the total sum of the premiums is less than the expected amount that is needed. The reason for this is that invested money can grow with a risk-free interest rate of $r\%$, which means an amount of money today is worth more than that same amount of money a year later. Using this information, it is possible to calculate the value of an amount money $Y$ received $n$ years from now as if it existed today. This process is called discounting. For example,

$$Y_{\text{present}} = \frac{Y}{(1 + r)^n}$$

$Y_{\text{present}}$ is the discounted value of $Y$ over $n$ years calculated with an interest rate $r$. $Y_{\text{present}}$ is called the present value. Calculations with present values are used to compare future cash flows with present ones. Using discounting, the value of every single contribution can be discounted to the present value. For example, imagine a person of 25 years old who wants to receive a pension contribution of €1 each year of his retirement. This means he still has 42 years to save up money for this. So at age 25 his invested money can still grow for 42 years.

When the probabilities that a person is still alive at a certain age are known, the yearly premium can be calculated. The costs for the insurer of €1 of pension is calculated by summing the present value of all contributions of €1 multiplied with the probability they actually have to be paid (i.e. the person is still alive). This results in:

$$\text{Premium}_{25} = 42|p_{25} \ast \frac{1}{(1 + r)^{42}} + 43|p_{25} \ast \frac{1}{(1 + r)^{43}} + 44|p_{25} \ast \frac{1}{(1 + r)^{44}} + \ldots$$

$$= \sum_{t=42}^{\omega-25} \frac{1}{(1 + r)^t} |p_{25}$$

(2.1)
where $t|p_x$ is the probability of a $x$-aged person reaching age $x + t$ and $\omega$ is the maximum age that people get, which will be discussed in the next section. $t|p_x$ is calculated by multiplying $1|p_x \times 1|p_{x+1} \times \ldots \times 1|p_{x+t}$. The probability a person of age $x$ reaching $x + 1$ is called the survival rate and the probability a person of age $x$ not reaching $x + 1$ is called the mortality rate, denoted by $1|q_x$. Naturally, the following holds: $1|q_x = 1 - 1|p_x$. For simplicity, the following notation is used: $1|p_x = p_x$ and $1|q_x = q_x$.

The sum of equation 2.1 can be calculated for every age $x$. Notice that the interest rate $r$ will be different every year. Insurers however, do their calculations based on the assumption of a constant rate. The calculated premium is the premium for €1 of pension. Paying the premium for age 26 means the person will receive an additional €1 on his pension. So the yearly contributions will become €2 per year. Notice that at low ages, the money can be invested many years, while at high ages the money can be only invested a few years. So at every age it will costs a different amount of money to build up €1 of pension. This results in a different premium for every age $x$.

Logically, the yearly pension contributions normally are higher than €1. To calculate the premiums for building up pension with €500 instead of €1, simply multiply the premiums found for €1 with 500. So the calculated premiums for €1 are essentially the base premiums. Throughout this report, these base premiums will be called insurer fees. The focus in this report will be on the insurer fees, since the different premiums can be derived from these fees.

### 2.2 Discount rate

Since the insurers cannot constantly account for the changes in interest, the fees are calculated under the assumption of a fixed rate. This fixed rate is based on the expected return on investments. A high return on investments results in a higher growth rate of the money that is set aside for pension. This means that when the expected return on investments is high, the insurer fees are lower.

DNB (De Nederlandse Bank) has formulated regulations regarding the minimum discount rate. They determine a risk-free rate of return on investments. These risk-free rates of return on investments is a guaranteed interest rate that insured persons must receive on their money, they have "saved up" with their insurer.

As seen in section 2.1, the insurer fees are based on the summation of the discounted expected payouts. DNB determines the discount rates for every yearly payout. These discount rates are determined by means of creditworthy obligations. For example, the payout on age 70 for a 20-year old person now is discounted with the 50-year discount rate of DNB. The payout at age 71 is discounted with the 51-year discount rate of DNB and so on. Therefore, every payout has its own discount rate.

Insurers also do these calculations precisely since they want to have insight on their expenses. However, for the simplicity of their customer, they calculate with a fixed discount rate based on the DNB discount rates. This rate changes along with the changes of the risk-free rates of DNB. DNB adjusts their rates according to the economic situation. The following figure shows the discount rate for investments of different durations in the period 2001-2012.
Figure 2.1: Risk-free interest rates for 1-year (solid), 10-year (dashed) and 30-year (dotted) investments in the period 2001-2012. Source: DNB (De Nederlandse Bank).

Figure 2.1 shows the risk-free interest rates according to DNB (De Nederlandse Bank). These rates represent the expected return per year for 'risk-free' investments of different durations. Examples of risk-free investments are creditworthy government bonds. It also shows that these discount rates change every year. In 2008 the rates dropped as a result of the crisis. This means that the insurers also have to adjust their discount rate and therefore their fees. Normally, this is not a problem for insurers, since they hedge their position with long term obligations. Changes in interest rate do not have an affect on these long term obligations. This means that the insurer has enough money to cover the payouts in the future. In new contracts with employers they do have to adjust their discount rate to the average DNB rate. They will then again hedge their future expenses with long term obligations. So essentially their risk is always covered.

However, this was not the case during the recession that started in 2008. What went wrong in the crisis is that many insurers did not hedge their interest risk fully. Therefore, money was lost and future obligatory payouts could not be fulfilled. As a result, pensions were depreciated.

Interest rates have a big impact on the pensions costs. Since insurers use a rate that is based on the DNB risk-free rates, it would be very interesting to see how DNB calculates them. But since this is a complete study on itself, we will not discuss it in this report. However, for more information about interest rate modelling, see [10].

Besides the risk-free interest rates, the insurer fees are also based on the mortality rates. Historical data of the mortality rates is available and we will examine this data in the next section.
3 Examination of the data

The previous section shows that the insurer fees are based on the interest rates, the mortality rates and the maximum age. The mortality rates have been changing constantly in history. It is therefore interesting to use the available information to investigate these developments. The data which is used in this project, is supplied by CBS (Central Bureau of Statistics). It contains the mortality rates for an \( x \)-aged person \((x = 0, \ldots, 100)\) in the years 1950, \ldots, 2012. Since these mortality rates are used in the models we are going to investigate, we will also use this data throughout our report. However, the uncertainty in these mortality rates will also be examined by looking at the actual proportion of amount of deaths and people alive at a given age.

3.1 Mortality rates

It is interesting to examine how the mortality has developed during the period 1950-2012. The mortality rates are determined by the number of deaths that occur at every age divided by the total population of that specific age. Logically, these numbers will differ for every year.

![Mortality rates at every age for men in different years. The years 1950 (dotted), 1990 (dashed) en 2012 (solid) are shown.](image)

Figure 3.1 shows that the mortality rates for men seem quite constant in the ages 5, \ldots, 50. After age 50 the rates increase exponentially. It is clear that this growth started at an earlier age in 1950 than 2012. The same patterns hold for the mortality rates of women.
This means that people are dying at a higher age. To make this becomes more clear, a population group of 100,000 persons is examined in different years. This population group is exposed to the mortality rates at every age of the corresponding year. The result is shown in the following figure.

![Figure 3.2](image.png)

**Figure 3.2**: Number of deaths for a male population of 100,000 at every age in different periods. The periods that are shown are 1950 (solid), 1990 (dotted) and 2012 (dashed).

Figure 3.2 shows the number of deaths in the years 1950, 1990 and 2012. Note that the decrease in number of deaths after age 80 is due to the fact that less people of the population group are alive at that point. This does not mean that the mortality rates are decreasing at that point. The figure also clearly shows that the infant mortality has decreased drastically. Another change is that the ages where the amount of deaths are the highest has shifted to a later stage, but has been more concentrated. This results in a big increase of people living past age 70, but almost no difference in people aged 90+. Again, the same patterns hold for women.

### 3.2 Maximum age

As already stated, people are growing older. The age of death is becoming more concentrated, however this does not immediately indicate that the maximum age of people has also grown. As the figure shows the age of death is shifting to the right, but it is also becoming more concentrated. The following figure shows the maximum ages reached in the past 50 years.
Figure 3.3 shows that the maximum has been growing in the past 50 years. A linear trend is fitted to the data. This linear fit indicates that the maximum age has grown with 5 years over the past 55 years.

Regarding the calculation of the insurer fees, the assumption of a maximum age of 120 was made in Equation 2.1. Since a linear fit is found for the maximum age, it is neater to let the maximum age be dependent of time. In retrospect, this means that the life expectancy in 1960 should be calculated with $\omega = 107$ and in 2012 using $\omega = 112$. Instead of using those $\omega$’s, an upper bound of $\omega = 120$ is used. However, this does not significantly influence the insurer fees. The reason for this is that the mortality rates are very high in ages 110-120, which results in a very low survival rate. Together with discounting, the affect that the high ages have on the insurer fees are neglectible.

For illustration, consider the calculation of the insurer fees for a 25 year old person:

$$ Premium_{45} = \sum_{t=42}^{\omega-25} \frac{1}{(1+r)^t} t|p_{25} \]

In 2012, the maximum age was 112 according to the linear fit in figure 3.3. Therefore, the fee is calculated with $\omega = 112$ and compared to the fee calculated using $\omega = 120$. Using the mortality rates in 2012 and a discount rate of 3%, this results in:

$$ Premium_{45} = \sum_{t=42}^{120-25} \frac{1}{(1+r)^t} t|p_{25} = 3.135923 $$

$$ Premium_{45} = \sum_{t=42}^{112-25} \frac{1}{(1+r)^t} t|p_{25} = 3.135923 $$

So there is no notable difference when using a different maximum age. For further illustration, the amount of money that every possible payout contributes to the fee is shown in the figure below.
The total fee of 3.135923 is the sum of the bars at every age in figure 3.4. So every bar represents a term of the sum given above. The values decrease because of a decreasing survival rate. Since the probability of reaching age 105+ is very low, the probability that an insurer has to do a payment is low. Therefore, the value decreases. Besides that, the payments at these high ages are also affected by many years of discounting.

After age 105 the bars are practically zero. This means that choosing a higher $\omega$ than the maximum age, does not influence the insurer fees. Choosing a $\omega$ that is lower than the maximum age does have an impact on the fee. It will result in a lower fee since bars greater than zero will not be taken into account. Therefore, insurance companies overestimate the maximum age when calculating the insurer fees.

As stated before, the insurers use the forecasted mortality rates to calculate the insurer fees instead of the historical ones. This means that the maximum age will also have grown. It needs to be checked whether the maximum age will grow past 120.

However, according to the linear trend in figure 3.3 the maximum age will have reached 117 in 2062. Since we are only forecasting up until 2062, this means that calculations can safely be done using the fixed maximum age of 120.

### 3.3 Uncertainty in the data

Intuitively, the mortality rates are determined by the amount of deaths at a specific age divided by the amount of people that were still alive. This data is also available from the CBS database. However, calculating this proportion in a certain year will only approximate the real probability, as this is only one sample. Logically, the amount of people alive decreases as age increases. This means that the mortality rates at high ages are calculated with a smaller population, which results in a higher uncertainty in the mortality rates at high ages than at low ages. The standard error of the mortality rates is calculated and a 95% confidence interval is given below.
Figure 3.5: Confidence interval of mortality rates for men in 2012 for ages 0-60 (above) and ages 60-100 (below). The figures are scaled differently.

Figure 3.5 shows the 95% confidence interval for the mortality rates. Please note that the figures are scaled differently. The uncertainty in mortality rate is the highest at age 100. This is the result of few people growing this old. Therefore, less data is available, which results in a higher uncertainty.
4 — Life expectancy

When the mortality or survival rates are known, the life expectancy of a person of age \(x\) is easily calculated. By definition the life expectancy is the amount of years a person of age \(x\) lives on. That is,

\[
y_{\text{remaining}} = \sum_{t=1}^{\omega-x} t \times t|p_x
\]

With this calculation the life expectancy of a \(x\) year old person becomes: \(x + y_{\text{remaining}}\).

Note that the life expectancy is not actually needed to calculate the insurer fees. As Equation 2.1 shows, the premium does only depend on the mortality rates and the discount rate. However, developments in life expectancy give more perspective than the developments in the mortality rates. If the life expectancy has grown with one year, this means that the insurers have one more expected payout, which immediately results in a higher premium. Changes in mortality rates are not that insightful, since the different mortality rates at different ages each have their own effect on the premium.

4.1 Developments in life expectancy

With the mortality rates the life expectancy can be calculated. It is interesting to see how the life expectancy has developed since 1950. In this report the amount of years between retirement and death will play an important role. Since 67 is the retirement age in the Netherlands, the life expectancy of a 67-aged person is investigated. As already stated, there exists some uncertainty in the mortality rates provided by CBS. Logically, this will also affect the life expectancy. To give a good indication on how accurate the calculated life expectancies are, a 95 % confidence interval is calculated.
Figure 4.1: Life expectancy men (black) and women (grey) at age 67 in the period 1950-2012. The 95% confidence interval is given by the dashed lines.

Figure 4.1 shows the 95% confidence interval of the life expectancies. The difference between the upper and lower bound fluctuates between 0.5 and 0.8 years for both men and women. This shows that the uncertainty in the mortality rates also affects the life expectancy. Luckily enough, the differences are not very significant, which makes our results pretty reliable.
5 — AG model

In the previous section the development of the mortality rates and life expectancies from 1950 up until 2012 is shown. Examination of these developments makes it very likely that the life expectancy in the future will probably differ from the current life expectancy. Since insurers are reserving money for payments of employees in the future, this means that the future life expectancies are important for the insurers. It is for this reason that the future mortality rates are predicted.

These mortality rates depend on a lot of uncertain factors, such as: medicines against diseases, possible epidemic outbreaks or war. Due to the presence of these uncertainties, it is difficult to give a precise prediction. Several different models exist for predicting the mortality rates, each one with its own approach. Some models take into account a lot of different factors, while others keep it relatively simple. One of the more commonly used is the AG-model, which belongs to this second category of models.

5.1 Approach of the model

The AG-model is built by extrapolating the historical data of CBS. To filter out noise, the historical data is smoothed using a smoothing algorithm. Since AG is a renowned institution in the area of insurances, the model will be examined thoroughly and compared to alternative models later in this report.

5.1.1 Definitions

The following terminology is used in the model.

\[
\begin{align*}
x &= \text{age} \\
t &= \text{year } = 1989, \ldots, 2012 \\
s &= \{m, v\} \text{ where } m = \text{male and } v= \text{female} \\
q_{x,s}(t) &= \text{mortality rate of a } x\text{-aged male or female in year } t
\end{align*}
\]

5.2 Smoothing

In Figure [3.1] the mortality rates for men are shown. The rates may seem to grow according to a flowing line, except for the higher ages. At higher ages the rates are jerking up and down. However, a closer look shows that this is also the case for the rest of the graph. This is not clearly seen because the fluctuations are relatively small. Zooming in on the graph will show this more clearly.
Figure 5.1: Mortality rates for men (solid) and women (dashed) in 2012. The graph is split in two parts in order to remark the jerkiness of the rates.

Figure 5.1 shows how the mortality rates jump up and down. This is not desirable, since most of these fluctuations are noise. To filter out this noise, smoothing is used.

5.2.1 Van Broekhoven Algorithm

In the AG model the mortality rates are smoothed using the Van Broekhoven (VB) algorithm. This algorithm uses indicator kernel smoothing, after having transformed the data using the Gompertz transformation (see: [4],[7]). To determine the smoothed value of a datapoint, a number surrounding datapoints are used. This number depends on the chosen bandwidth. For every surrounding datapoint that is in the range of the bandwidth, a weight is determined. These weights are found in the following formula.

\[
f(qG_{x,s}) = \sum_{k=-m}^{m} \left( \frac{12m^2 + 24m + 5 - 20k^2}{32m(m+1)(m+2)} \right) qG_{x+k,s}
\]

where \(qG_{x,s}\) are the Gompertz transformed mortality rates and \(m\) is the chosen bandwidth. This results in smoothed transformed mortality rates. Using the inverse transformation on these smoothed rates, results in the desired normal smoothed mortality rates. See appendix A for the determination of the weights and a detailed explanation of the algorithm.

For the VB algorithm a bandwidth \(m\) must be chosen. AG choice for this bandwidth is \(m = 4\). The choice of \(m\) is very important for the algorithm. A high \(m\) results in a smooth estimation, while a small \(m\) does not. However the smoothed values correspond better to the real values when a small \(m\) is used. It is interesting to examine how the smoothed rates differ for different \(m\). For comparison, the rates for men in 2012 are shown.
Since the smoothing is very accurate up to age 60, only the ages 60 to 95 are shown in figure 5.2. The figure clearly shows that $m = 5$ results in a pretty accurate estimation of the real mortality rates. For small $m$ the mortality rates are underestimated, while large $m$ overestimate the data.

To give an indication of how well the smoothed rates fit the original mortality rates, the sum of squared errors (SSE) is calculated for different $m$. Logically, a small SSE is desirable. It is interesting to look at how the SSE changes when a different $m$ is used.

When comparing the SSE for different $m$, a problem can occur on the edges. Since there are not many datapoints on the edges, a large $m$ cannot be used. Therefore, $m$ must be adjusted on the edges. For example, consider the mortality rate at age $x = 1$ when trying to smooth with $m = 3$. This means the mortality rates for ages $-2, -1, 0, \ldots, 4$ must be used. Obviously, no mortality rates exist for negative ages. Therefore, $m$ must be adjusted at these low ages. In this example, the $m$ must be adjusted to 1 in order to have no negative ages. The same problem holds for high ages, where no data points exist for ages 100+. In this case the same solution is used, where $m$ is chosen smaller, such that no non-existing values are needed. When this is done properly, the SSE’s can be compared fairly.
As figure 5.3 shows, different $m$’s result in a different SSE, as expected. But increasing $m$ does not simply result in a smaller SSE. The minimum value for the SSE occurs when $m = 4$ and $m = 5$, with minor differences. The use of $m = 4$ by AG therefore is a valid choice.

Again, these comparisons were made based on the mortality rates of men in the year 2012. When the mortality rates for women or the rates at different periods are examined, the same patterns hold.

5.2.2 Exponential smoothing

In section 5.2.1 the van Broekhoven Algorithm was used to smooth the mortality rates. Other approaches also exist, for example finding a function which fits the data.

Examination of the mortality rates shows that there is no known function, which fits the mortality rates. However, if the rates are divided into parts, different functions can be found for every part. Observation shows that in early ages (e.g. 0-10) the rates behave like a decreasing exponential function. The middle ages can be described using a linear function and the ages $> 60$ behave very much like a increasing exponential function. However, it is not clear at what age the decreasing exponential turns into a linear function, or where the linear turns into an increasing exponential. To determine these ages the following method is used.

First, a decreasing exponential fit is determined for ages $x = 0 \ldots x^*_t$. Then a linear function is fitted to the data for ages $x^*_t \ldots y^*_t$ and lastly, an increasing exponential function is fitted to ages $y^*_t \ldots 100$. All of these fits are determined using OLS (ordinary least squares). The sum of the squared error of the three fits is calculated for all possibilities of $x^*_t$ and $y^*_t$. The best fitting $x^*_t, y^*_t$ are now found by choosing the ages where the sum of the errors is lowest. This method
can be used for every period $t = 1950 \ldots 2012$. For men in 2012, the following results are found:

\[
\begin{align*}
x_{2012}^* &= 5 \\
y_{2012}^* &= 31
\end{align*}
\]

So this means that a decreasing exponential function is fitted to ages 0-5, a linear function to ages 6-31 and a positive exponential to ages 32-100. This method is compared to the van Broekhoven algorithm.

Figure 5.4: The goodness of VB (dashed) and function fits (dotted) in approximating the actual mortality rates (solid). The data is split up into ages 0-50 (above) and ages 51-100 (below) for a better insight.

Figure 5.4 shows that the use of different functions results in a good estimation of the known mortality rates. To accurately compare the both smoothing methods, the total errors are compared.
### Table 5.1: Error of the fits

<table>
<thead>
<tr>
<th>Method</th>
<th>Total squared error</th>
</tr>
</thead>
<tbody>
<tr>
<td>VB</td>
<td>0.00407</td>
</tr>
<tr>
<td>Functions</td>
<td>0.00289</td>
</tr>
</tbody>
</table>

This means that the function fit method gives a 1.4 times more accurate estimation than the VB algorithm. Examination of other years results in different errors for both methods. Comparison of these errors shows that both methods perform equally well. The function fit method performs better than the VB algorithm in exactly 50% of the years. This means that both methods are equally accurate in smoothing the mortality rates. Therefore, the strengths and weaknesses of both methods are discussed.

The VB algorithm does not respond well to the exponential growth of the mortality rate, it does however take the small changes into account. The function fit technique results in a very accurate estimation at ages 60+ since the growth is described well by an exponential function. The downside however, is that the technique does not respond well to small increases in growth. For example, the growth in mortality rates due to the 'accident hump'. This hump is caused by careless behaviour of men between ages 16 and 26, which results in a small growth in mortality. Besides that, the calculation of the ideal $x^*$ and $y^*$ and determining corresponding fits is extensive. This is because of the fact that for every possible option, the SSE is calculated to determine the minimum. For this reason the method is not elaborated any further in this report and VB will be used for smoothing the mortality rates.

### 5.3 Mortality rates for high ages

There is not much useable data for the mortality rates in high ages (100+). This data is not available since there are not many people who survive until these high ages. The oldest person ever in the Netherlands reached an age of 116. When there are so few people the mortality rates estimates are very inaccurate and can not be used properly.

In retrospect to section 5.2.2 there is a reason to believe that the high ages mortality rates can be described using an exponential function. The function fit technique performed well on the ages 60+, where a exponential function was fitted to the mortality rates. However, using this exponential function at the ages 100+ will result in mortality rates of 1 at age 105. Logically, this is not a desirable result since people can grow older than age 105. Therefore, a model must be used in which the growth decreases.

Many theories how the mortality rates behave at these high ages and how they can be modelled. Some theories believe the growth of mortality rates will decrease and reach an asymptotic level, while other theories state that they will continue to grow to a probability of 1. These theories are discussed in the appendix B. AG uses the logistic mortality law of Kannisto (see: [7]), which states that the growth of mortality will reach an asymptotic level. This will result in the following mortality rates for men and women.
Figure 5.5 clearly shows that the growth of the mortality decreases in the ages 105-120. With the smoothed mortality rates and the modelled mortality rates for high ages, a forecast of the mortality rates in the next 50 years can be done.

5.4 The forecasting model

The smoothed mortality rates of the periods 1950-2012 can be used to forecast the future mortality rates up to 2062. This is done by extrapolating the smoothed mortality rates for ages 0,...,120. As seen before, the rates per age change every year. Mostly, the mortality rates have been decreasing. The rate of this decrease however, differs from year to year.

When the rates of decrease are investigated, some patterns can be found. For example, if the rate of decrease increases with 1% every year for 5 years, a 5 year pattern is found. These are the trends in development of the mortality rates.

In the forecasting model of AG a short term trend and a long term trend is calculated. The long term trend will determine the mortality rates in 2062, where the short term trend describes how the mortality rates behave between 2012-2062.

5.4.1 Long and short term trends

The choice of the long and short term trend is based on the log mortality rates in the years 1950,...,2012. By taking the log of the mortality rates, trends can be spotted easier. In the model 9 and 23 years are used for respectively the short and long term trend. It is interesting to examine why AG has chosen for these specific amounts of years.
Figure 5.6: Development of the log-mortality rates in the period 1950 to 2012. The rates are shown as a fraction of the rates in 2012.

Figure 5.6 shows the mean log-mortality rates in every year as a fraction of the mean log-mortality rate in 2012. The average mortality rate over every age is shown. This means that only a general trend will be shown. It would be a lot more informative if the mortality rate at every age could be examined separately. However, this will result in 100 figures and therefore is not a feasible option.

The log mortality rates in Figure 5.6 are increasing, which means that the actual mortality rates are decreasing. The figure not only shows that the mortality rates are decreasing, but also that they are decreasing more rapidly. Some trends can already be spotted by observation. For example, in the period 2004-2012, there has been a somewhat constant growth in the log mortality rates. This period of 9 years is therefore a logical choice for the short term trend. If this trend also holds for the coming years is uncertain. It is possible that this growth will stagnate, or it is even possible for the log-mortality rates to decrease (and therefore for the mortality rates to increase). To find the fitting trends, linear functions are fitted to the data. This will help in determining the short and long term trends that are needed in the forecast.
As Figure 5.7 shows the 19-year fit does not represent the log mortality rates very well, where the 7-year fit does. Remember that this means there is a constant average decrease in mortality rates in the period 2005-2012. So as a choice for a trend, the 7-year trend is a valid one.

It is interesting to compare the mean squared error of the fits. This way the goodness of the fits can be compared fairly. The figure below shows the mean squared error for different durations of trends.

Figure 5.8 shows that the MSE increases when the duration of the trends is longer. The
short year trends have a smaller error, which is logical because there are not many points that have to be fitted.

However, the problem that arises is that it is not clear whether a trend is a short or a long term trend. Since the choice of a one-year or two-year trend leads to a minimum error, that might seem like the optimal choice. However it is intuitively clear that these choices are not representative for the future estimation. But it is also not clear whether a 6-year trend is an appropriate choice. And at what point does a short term trend turn into a long term trend? This is a somewhat grey area where there is no correct answer. Therefore, different choices for trends will be examined and an alternative approach will be used in section 5.4.3.

5.4.2 Forecasting the mortality rates

As stated before, the choice of AG is 9 years as a short term trend and 23 years for the long term trend. As figure 5.8 shows, these choices are very decent. With these choices, it is possible now to forecast the future mortality rates. This is done as follows:

\[ q_{x,s}(t) = q_{x,s}^{start} \left( f_{short}(t) \right) t_{hor} \]

This formula is constructed as follows: First the \( f_{short} \) and \( f_{long} \) are calculated as the geometric average of the mortality rates for respectively 9 and 23 years. So for the AG model these are as follows:

\[ f_{short} = \left( \frac{q_{x,s}(2012)}{q_{x,s}(2003)} \right)^{\frac{1}{9}} \]

\[ f_{long} = \left( \frac{q_{x,s}(2012)}{q_{x,s}(1989)} \right)^{\frac{1}{23}} \]

The mortality rates in 2062, also called the goaltable, are calculated as follows:

\[ q_{x,s}^{goal} = q_{x,s}(2012) f_{long}^{t_{hor}} \]

where \( t_{hor} = 50 \) is the amount of years that are forecasted.

The next step is the actual forecasting method. Because of the fact that the mortality rates are decreasing to 0, the assumption is made that the decrease in mortality rates will slow down. AG assumes this can be described using an exponential. This results in:

\[ q_{x,s}(t) = q_{x,s}^{start} \prod_{j=1}^{t} f_{short} e^{\alpha_j} \]

\[ = q_{x,s}^{start} t_{hor} e^{\frac{t_{hor}+1}{2}} e^{\alpha x} \]

In this expression, \( e^{\alpha x} \) will supply the slower decrease in mortality rates. \( \alpha x \) can be found by equating \( q_{x,s}(t_{hor}) \) with \( q_{x,s}^{goal} \).

This results in:

\[ \alpha x = \frac{\log(q_{x,s}^{goal}) - \log(q_{x,s}^{start}) - t_{hor} \log(f_{short})}{t_{hor}(t_{hor}+1)} \]

Inserting this expression into equation 5.2 will result in equation 5.1.
As the mortality rates at each age are forecasted, the life expectancy can be calculated. Regarding pensions, it is interesting to look at life expectancy given age 67 is reached. This shows the expected amount of years a person receives pension contribution.

Figure 5.9: Forecast life expectancy of a 67-year old men (above) or women (below) for different trends. The symbols indicate the life expectancy in 2062 for a long term trend of 16 (circle), 20 (dot), 29 (triangle) and 40 years (diamond). For short term trends 6 (dashed) and 15 years (dotted) are illustrated.

Figure 5.9 shows the forecasted life expectancies up to the year 2062 for a 67 year old man or woman. There are graphs for several different combinations of long and short term trends. Up to 2030 a pretty consistent life expectancy is found, but during the period 2030-2062 the
projections drift apart. Choosing different long term trends results in a different life expectancy in 2062. So the long term trend determines how much the life expectancy will grow in 50 years. The short term trend determines how the life expectancy grows towards this life expectancy in 2062.

For men, choosing a longer long term trend results in a lower life expectancy in 2062. For women however, the life expectancy is lowest when the long term trends is around 30 years. The reason for this can be derived from the fact that the life expectancy of a 67-year old woman did not grow in the period 1985-1995 (see Figure 4.1). So taking this period into account for the choice of the long term trend will result in a small growth of the life expectancy.

Figure 5.9 shows a maximum difference of 2.7 years. A very considerable difference, since this is only the result of choosing different trends. This means that the choice for trends is very important. However, since there is no clear choice for what trends to use, it is hard to say whether or not a good forecast is done. Therefore, an alternative method is examined.

5.4.3 Weighted term forecast

As there is no certainty of what trends will hold in the future, choosing just two trends is a risky approach. It is also possible to take every trend into account in the forecast, having more or less influence depending on how well they fit. To determine the goodness of fit of the trends, a reflection to figure 5.8 is useful. This figure shows the errors for the different linear fits that can be fitted to the growth of the mortality rates (see figure 5.7). These errors can be used as a weight for the projection. Using this, the mortality rates can be constructed with weighted forecasts. Trends with big errors will receive less weight than trends with a small error.

So the method is as follows: First, the linear fits of different periods for the short term trends are determined together with their errors. The errors indicate how much weight a period will receive. For example, period 2002-2012 has a relatively high error, therefore it receives less weight than period 2006-2012, which has a smaller error. The total sum of these weights is 1. The same method is used to determine the weights for different long term trends. Secondly, the forecast of the life expectancy is calculated for every combination of short and long term trends. Some of these combinations are also shown in figure 5.9. After that, the estimated life expectancies are multiplied with their corresponding weights for long and short term trends. Summation of the life expectancies for both long and short trends results in a weighted trend life expectancy forecast. This method is used for both men and women. Since the rate of decrease for the mortality rates of women is different than men, the best fitting trends and therefore the weights will also differ.
Figure 5.10: Forecasted life expectancies using weighted trends (dashed) compared to AG’s choice for trends for both men (black) and women (grey).

Figure 5.10 shows the life expectancy forecast using the trends of AG’s choice, namely $\text{years}_{\text{short}} = 9$, $\text{years}_{\text{long}} = 23$. Besides that, it also shows the weighted trends forecast. The life expectancy for women grows a little bit larger when weighted trends are used. The difference in 2062 is approximately 0.5 years. The benefit of the weighted trend method is that every possible choice for trends is taken into account. Depending on how well they fit the data, they have less or more influence. Besides that, a distinction can be made in the trend for men and women. This is a benefit because historical data shows the behaviour of life expectancy of men and women is notably different.

Therefore, we conclude that the weighted trend forecast could be considered as an addition to the AG-model.
In the previous section the forecasting model of AG has been examined. The assumption was made that the decrease in mortality rates slowed as the years increased. This resulted in a rapidly growing life expectancy in the first 25 years of the forecast. This growth however, decreased in the last 25 years of the forecast.

Another model to forecast the future life expectancy is the model of the PLT (pensioen en lijfrente-tafel) workgroup. The approach of the PLT workgroup does not differ much from the AG model. The development of the mortality rates of the period 1962-2012 is examined. Using this information, the rate of decrease for every age in every period is used to forecast the mortality rates up to the year 2062. Since it has a similar approach as AG, this chapter will only describe the model briefly. For a detailed description see [3].

6.1 Reduction factors

In the period 1962-2012 a decrease in the mortality rate is noted. The reduction factors \(a_{x,t}\) are the factors of how the rates decrease. In this model the average reduction factors are calculated regarding to the year 2012. For example, if the mortality rate at age 80 is 0.2 in the year 2007 and this rate decreases to 0.15 in 2012, the reduction factor is \(\frac{0.20}{0.15} = 1.33\). This means that in the period 2007-2012, the mortality rate at age 80 have decreased with an average factor of 1.06 per year.

Using the reduction factors of the period 1962-2012, future reduction factors are forecasted for the years 2013-2062. The assumption is made that the average reduction per year in the future will be the same as it was in the past. Calculating the factors goes as follows:

\[
a_{x,t} = \left( \frac{q_{x,2012}}{q_{x,2012-k^*}} \right)^{\frac{1}{k^*}}
\]

where \(k^* = \max(t - 2012, 5)\).

When looking back at the example, this calculation will result in a expected reduction factor of 1.06 at age 80 in the year 2017. A minimum of 5 years is used because the mortality rates can fluctuate heavily in 5 years. Therefore, PLT uses a minimum of 5 years because they believe changes over a 5-year period are more substantial.

After the factors are calculated, they are smoothed for every period by taking the mean of 11 ages surrounding \(a_{x,t}\). See appendix A.3.2 for an explanation of why 11 factors are chosen. The smoothed factors are denoted as \(\overline{a}_{x,t}\) and are calculated as follows:

\[
\overline{a}_{x,t} = \frac{a_{x-5,t} + a_{x-4,t} + \ldots + a_{x,t} + \ldots + a_{x+5}}{11}
\]

This is not possible for ages: \(x = 0,\ldots,4\) and \(x = 90,\ldots,94\). At these ages the maximum amount of surrounding ages with an equal amount on both sides is used.

The mortality rates for high ages are estimated the same way as the AG-model. Namely, using Kannistō’s model.
6.2 Forecasted mortality rates

Using the expected reduction factors, the future mortality rates can be forecasted. They are calculated recursively for every age $x$ using the formula:

$$q_{x,2012+k} = \pi_{x,2012+k} \cdot q_{x,2012+k-1}, \quad k = 0, 1, \ldots, 50$$

These mortality rates are translated into life expectancies using the same method as before. This results in different life expectancies than the AG-model forecast. The following figure illustrates the difference.

![Figure 6.1: Forecasted life expectancy of 67-aged men (black) and women (grey) using the PLT-model (solid) compared to the results of the AG-model (dashed)).](image)

As the figure clearly shows the PLT forecast has a more optimistic life expectancy than the AG forecast. The average difference in 2062 between the two models is approximately 2.8 years. This difference can be explained by the assumptions of both models. AG assumed that the decrease in mortality rates will slow down, whereas PLT expect them to behave exactly like the past 50 years. This results in lower forecasted mortality rates in the PLT-model and therefore a higher life expectancy.
7 — Performance analysis

In the previous sections, the models of AG and PLT were investigated. Both models use extrapolation of historical data; however, different results are found. Since the future life expectancy is not known, the performance of the models can not be validated. This means that it is hard to check which model is best. This will only be known afterwards.

However, the models can be used on less data, such as the period 1950-1995, to give a prediction of the known life expectancies in 1996-2012. By using 1995 as the starting year of the prediction, 45 years of data can be used. Since this is a lot of data, the models should do a good job in forecasting the period 1996-2012. Using this method, the models can be rated on how well they would have performed in the past. This gives some indication on how well they will perform in the future.

Figure 7.1 shows that the life expectancies of both men and women is modelled well up to 2003, that is 8 years. The PLT-model is a bit more accurate than the AG-model for both and women. After 2003, the life expectancies start to grow faster. Both models would not have predicted this growth according to the data of 1950-1995.

Of course it is also possible to use any period other than 1950-1995. This shows that the performance of both models is not very good. Mostly, the models’ predictions do not hold for more than 4-5 years. The PLT model does performs a little bit better than the AG model.

Especially during the past 10 years, the performance of both models is poor. This is because of the big increase in life expectancy in the period 2003-2012, that both models do not predict when data 1950-2003 is used. This shows that the models are not good for predicting the life expectancy many years ahead. It is probably for this reason that the models are updated every two years. This means that every two years the life expectancy prediction is adjusted according to the new available data.

Because insurers have contracts, they can not change their fees every two years. Most insurance companies have 5-year contracts with their clients. This means that after 5 years they have the possibility to adjust their fees according to the new models. This means that the performance of the models is sufficient for insurers most of the times. The forecasts are pretty accurate for 5 years, which means that the corresponding fees are substainable.
Figure 7.1: Forecast life expectancy of a 67-year old men (above) or women (below) based on the data of 1950-1995 (circles). The forecast is done using the AG-model (solid) and PLT-model(dashed). The triangles represent the calculated life expectancy of 1996-2012.
Comparison and construction of insurer fees

The previous sections have shown how mortality rates, life expectancy and interest rates behave. Based on this information, insurers construct their fees, such that all expected future expenses are covered. Different fees are constructed for man and women in the ages 15-67. Insurers calculate their own fees with their own method, which means that different insurers use different fees. By law, the fees an insurer uses, must be the same for every client/company. In this section the fees of different insurers are compared. Besides that, the method to construct the fees is examined and compared to alternative approaches.

8.1 Insurer fees

Insurers have different ways to determine their fees. Some insurers simply use the calculations of AG or PLT, while others do their own projections for mortality rates. In this section the fees of different insurers are compared with each other. The insurers that are examined are: Nationale Nederlanden (NN), DeltaLloyd (DL) and AEGON.

![Graph showing insurer fees for different insurers.](image)

AEGON and Nationale Nederlanden compare very well. As figure 8.1 shows the two lines are almost the same. Delta LLoyd is considerably higher than the other two. This is because...
they have used a lower discount rate to calculate their fees. Since this is a big difference, Delta Lloyd probably compensates this with discounts.

8.2 The use of forecasted fees

With the forecasted mortality rates at their disposal, insurers can produce their fees per age for every year up to 2060. Remember that the fees represent the costs that a $x$-year old person has to pay to receive €1 of pension. Since the mortality rates are different every year, every year also has different fees. These fees are called tables, such that the fees in 2012 are represented by the 2012 table. The question that arises is: Why and how do the insurers use the forecasted fees?

Companies have employees with different ages, this means that not all of the employees retire in the same year. For example, a person of age 25 will retire in 2056, while a person of 65 retires in 2016. Because the average life expectancy of a 67 year old in 2016 is significantly lower than the life expectancy in 2056, the pension costs will differ. This means that the estimated pension costs of the 25 year old is higher than the 65 year old. So the appropriate fees for the 25-year old are actually found in the 2056 table. In contrast, the 2016 table will represent the fees for the 65-year old person, since this is the year of his/her retirement. Taking into account these differences for every employee is a lot of work. Therefore, insurers averages out these differences and use a fixed table somewhere in the middle. This results in the choice of a table somewhere between 2035-2040, which differs per insurer. The insurers are legally obliged to use the table for every company and every employee in that company. This does not mean that every employee pays the same fee. As stated before, distinctions are made between gender and age.

The consequence of this for employees is illustrated in the following example: A 25 year old man has a pension contract (through his employer) with an insurer who uses the 2036 table. The man will retire in 2056, 42 years from now. Since 2036 table is used, he now pays less than the calculated price for him. 20 years later, in 2034 the man is 45 years old. The insurer will have chosen a different table. Assume this table is also shifted 20 years, which means the 2056 table is now used. This means the fees correspond perfectly with the pension costs of the man. However 20 years later, 2 years before his retirement, he will pay too much for his pension costs. Assuming the insurers table has also shifted with 20 years, the man will now pay the costs for a 65-year old who retires in 2076 instead of 2056. So the result is that an employee will pay more than the appropriate price for him when he is old, while this is compensated in his younger years.

8.2.1 Fee comparison

To calculate the fees according to both models, a table choice must be made. According to CBS, the average male employee’s age is 42.4, while for women this is 40.7. This means that the average female will work for 26.3 more years before retirement, while for men this is 24.6. Therefore, the forecasted fees are chosen that correspond to the year 2039 (2014+25) for men, and for women the fees that correspond to the year 2040 (2014+26). The resulting modelled fees are compared to the insurer fees.
Figure 8.2 shows that the calculated fees for both models are very similar. The fees of the insurers are a little bit higher than the modelled fees but it is a minor difference. The fact that both models have modelled very similar fees is interesting. This might seem wrong since the life expectancies are very different as figure 6.1 has shown. However, the calculated fees are based on the forecasted mortality rates in 2039 and 2040. In this period the forecasted life expectancy of both do not differ that much. The models are giving different results when later years are chosen.

The differences between the fees of both insurers and the models are relatively small when examining figure 8.2. However, the following example shows these differences more clearly.

Consider a 25, a 45 and a 65-year old male and female. These persons all have the same income of €50,000,-. They all build up €1000 of pension every year, which is 2% of €50,000,-. The following table shows the prices of the different insurers for these persons. Besides that, it also shows the costs when using the fees as calculated with the models.

<table>
<thead>
<tr>
<th>Insurer</th>
<th>Male 25</th>
<th>Male 45</th>
<th>Male 65</th>
<th>Female 25</th>
<th>Female 45</th>
<th>Female 65</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nat. Ned.</td>
<td>€4204</td>
<td>€7628</td>
<td>€14299</td>
<td>€4498</td>
<td>€8150</td>
<td>€15178</td>
</tr>
<tr>
<td>AEGON</td>
<td>€4152</td>
<td>€7533</td>
<td>€14136</td>
<td>€4290</td>
<td>€7777</td>
<td>€14755</td>
</tr>
<tr>
<td>Delta Lloyd</td>
<td>€5637</td>
<td>€9267</td>
<td>€15643</td>
<td>€5759</td>
<td>€9471</td>
<td>€16214</td>
</tr>
<tr>
<td>AG</td>
<td>€4031</td>
<td>€7330</td>
<td>€13910</td>
<td>€4238</td>
<td>€7694</td>
<td>€14695</td>
</tr>
<tr>
<td>PLT</td>
<td>€4078</td>
<td>€7424</td>
<td>€14134</td>
<td>€4191</td>
<td>€7623</td>
<td>€14648</td>
</tr>
</tbody>
</table>

Table 8.1: Costs for building up €1000 of pension at different insurer for different persons. The prices according to the PLT and AG model are also shown.
The table shows that minor differences in fees, can make big differences in total costs. The amounts in the table are only for one person. This means that for big companies, different fees can make an huge impact on the total employee’s costs.

8.3 Using alternative approaches

As the previous examples have shown, the calculation of the insurer fees is not very precise. It takes the mortality rates in a certain year and with these mortality rates, the fees are calculated. However, the mortality rates will not correspond to the actual rates a single person is exposed to. This means that wrong fees are calculated. Alternative approaches can be used to construct fees that match the individual person better.

8.3.1 Retirement tables

It is possible to calculate the costs for every employee when the table is used that corresponds to their retirement year. Using this method, every age will have its corresponding table, as in the previous example.

This method is compared to the 'standard' fixed table method in the following figures.

![Figure 8.3: Comparison of fees for men determined by using a fixed table (solid) of the year 2039, with the fees determined by using the corresponding retirement table (dashed). The retirement tables are calculated for both the AG (left) and the PLT model (right).](image)

As the figure shows the costs for a young male employee will be higher using this method, while the costs for an old employee is lower. The different models show similar patterns. The same pattern also hold for women. To illustrate this further the procentual differences per age are shown in the following figure.
Figure 8.4: Average procentual differences for men and women, while using of a different table for every employee compared to a fixed table. The upper figure shows the differences for the AG model, while the lower figure represents the PLT-model.

As figure 8.4 shows the use of the same table for different employees results in negative and positive differences, depending on the age. The young employee pays less, while the old employee pays too much. However as already stated, these differences will eventually cancel each other out.

Since old people pay less, using this technique would be very favorable for companies that have many old employees. However, it is legally obligatory for insurers to use the same table for
every employee. This means that this technique can not be used.

8.3.2 Diagonal table

It is also possible to use a technique to calculate the fees that is even more precise. The method uses the so-called diagonal table. It is based on the fact that, the mortality rates change every year.

Normally, the fees are calculated by summation of the discounted expected payouts. These expected payouts are calculated using the mortality rates at every age in a certain year. For example, a 25-year old person in 2014 is considered. Using the preceding methods the fee for this person is calculated using the mortality rates in a fixed year (2035 or 2056). To be more precise, mortality rates in different years should be used. In 2014 this person is exposed to the mortality rate of a 25-year old person. In 2015, the person will be exposed to the mortality rates of a 26-year old in 2015. Of course, this is a different rate than the mortality rate of a 26-year old in 2014. If this calculation is continued, the person will be exposed to the mortality rate of 67-year old in 2056 and eventually to the mortality rate of a 100-year old in 2089. Using this method the person is exposed to the corresponding forecasted mortality rates in each year. So therefore, this method is more precise. Note that, using this method the prices are actually calculated by looking at the mortality rates as a diagonal. Doing this for every age results in a new fee table. This table is denoted as the diagonal table. A further explanation is found in the appendix D.

The diagonal table is compared to the other approaches.

Figure 8.5 shows how the diagonal table compares to the other approaches. It is a bit more expensive than the retirement tables approach, however it shows the same behaviour of being cheaper for old employees.

The diagonal table has the same strength as the retirement tables, as it calculates fairer than the standard approach. It also has the benefit of being only one table, and therefore the same table can be used for all employees. This means that is legally allowed to use this table.

The problem with the diagonal is that although it is a more precise calculation, it relies on a larger forecast. For example, to calculate the fee for a 66-year old, a 54-year projection is needed (assuming a maximum age of 120). This amount increases as age decreases. Namely, for a 20-year old a 100-year projection is needed.

8.4 The benefits of alternative methods

The insurers use the same table for every employee, which is not entirely accurate as figure 8.5 shows. Their year of choice is determined by the average age of the employees. Insurers use this averaging technique because it is simple. Using a different table for every employee will cost a lot of extra administrative effort.

The differences for the old and young employees does not have a lot of impact on the eventual total costs for employees. However, since employers partially pay for the employee’s pension it is interesting for them to consider the alternative approaches. For many companies, the employees’ costs are a big expense. This means that a decrease of a few percents in costs is very beneficial. For a company with an average employee age higher than 42, the costs when using the diagonal table is lower than using the standard method. However, there are no insurers who make use of the diagonal table yet. This means that the first insurer who will introduce this system will be able to attract a lot of "old" companies. The other benefit is that the system is more transparent than the fixed table. The fees of the diagonal table are based on the future mortality rates. These rates correspond exactly to the age of the person in every year. So the calculation is 'fairer' than when a fixed table is used.
The downside of using the diagonal table is that this it will take a forecast of at least 100 years to calculate the fees at ages < 20. Compared to a fixed table in 2040, this means the forecast must be 3 times larger. Besides that, the maximum age must also must be forecasted. If the trend of the maximum age persists, the maximum age will be well over 120 in the year 2100. Therefore, a fixed maximum age of 120 can not be used in this calculation.
Figure 8.5: Comparison of fees determined by using a fixed table (solid), the fees determined by using the corresponding retirement table for every employee (dashed) and the fees calculated using the diagonal table. The upper figures represent the fees calculated using the AG-model, where the lower figures respresent the PLT-model. The figures on the left are the fees for men, where a fixed table of 2039, while the right figures show the fees for women with a fixed table of 2040.
A person in the Netherlands saves money for his pension during his working years. This is done according to a contract his employer has with an insurer. The insurer must make an estimation of how long the average person will live after his retirement. By looking at the life expectancy, the expected amount of payout years can be determined. Insurers calculate the life expectancies and therefore the expected future payouts, using forecasted mortality rates.

Different models exist to forecast these mortality rates. However, the available data is not perfect. There is uncertainty in the historical data. Besides that, the mortality rates for ages 100+ must be modelled because there is no useful data. The forecasts that are made, are based on trends of the past years. Trends are easy to spot, but there is no guarantee that a trend will persist. Since no medical developments are taken into account, it is hard to determine a lasting trend. It is probably because of this, that the models do not give an accurate forecast for more than 8-10 years, as seen in the performance analysis. It is therefore very sensible that the models are updated every 2 years.

The models investigated in this report are based on trends in the past years. Trends can change however, and since the models do not take uncertainty in the forecast into account, the results are discussable. A model that does take the uncertainty into account is the CBS model. This model is more complex than the ones of PLT and AG, and even based on medical developments and projections. It would therefore be very interesting to investigate. The complexity however, restrains us from investigating it in this report.

Insurer fees depend heavily on interest rates. As the money that is set aside can be invested at a certain risk-free rate, the amount of money would grow. This risk-free rate is called the discount rate. A high discount rate results in low fees, where a low rate results in higher fees. As a result to the crisis, discount rates have dropped, which resulted in higher fees. It should not have resulted in a shortage in pension funds however. This happened because the interest rate risk had not been covered completely, which resulted in the loss of a lot of money.

Fees are currently calculated by using the mortality rates in the year 2035-2040 (differs per insurer). This is because the average age of an employee is around 42, which means that the average retirement year is 25 years from now. This is not a neat calculation. If a person grows one year older, the year also has changed. To be completely accurate, the person must be exposed to the age dependent mortality rates that correspond with the year he will reach that age. To do this calculation, a forecast of 100 years is needed. As stated, the forecasts are only accurate for at most 10 years. So using this method will result in a precise calculation with inaccurate forecasts. It is therefore not an improvement in accuracy to use this method. However, it does result in a fairer calculation for the individual. Old people will pay less, where young people will pay more. This means that when an insurer would implement this alternative method, they will be financially attractive for relatively old companies.

Further investigation in this subject is definitely interesting. The investigated models are both based on extrapolation. This means that they are non-parametric models. This means that it is difficult to take the inaccuracy in the forecast into account. If it is taken into account and this is done correctly, an indication can be made on how accurate the forecast is.
Problems also exist in modelling high ages, both AG and PLT use Kannistö’s model for high ages, which presumes that late life mortality deceleration exists. Since there is few useful data about the high ages, it is hard to validate Kannistö’s model. A recent study has dismissed the deceleration and believes that Kannistö’s model is underestimating the mortality rates at high ages. Therefore, the developments in studies of late life mortality rate deceleration must be monitored. This way the model can be adjusted with an other high-age model when that proves to be necessary.

In this report the assumption was made that the maximum age will grow according to a linear trend. However, the possibility exists that the maximum age could behave like a polynomial or even an exponential. Since the maximum age is used in determining the fees, it is an interesting subject to investigate.

Investigation of the CBS model would be very helpful. Since the model investigates all causes of death, trends can be spotted more easily. This will make it possible to respond better to sudden increases or decreases in mortality. These trends can also be used in the AG and PLT model. The weighted term trend can also be considered, since it takes into account every possible trend.

The other benefit of the CBS-model is that it is a parametric model. This means that it also can easily take into account uncertainty in the forecast. With this uncertainty, a confidence interval for the fees can be constructed. This way, insurer have a better insight on the risk they take when choosing a certain fee.

The fees depend heavily on the discount rate. A small difference in the rate results in huge differences in the fees. Since the fees are so dependent on the discount rate, it would be nice if it could be predicted. Many interest rate models exist to do this, however these models also show how unpredictable the interest rates are. Since the fees are already based on uncertain forecasts of mortality rates, uncertainty of the interest rates will not be so helpful. However, it might be a possibility when implemented correctly.
Bibliography


[10] Roger J. Grabowski: developing the cost of equity mortality risk: Risk-free rate and ERP during periods of "flight to quality", 2011: ASA


A — Van Broekhoven Algorithm

The Van Broekhoven (VB) algorithm has been used extensively in life expectancy models over the past decennia. With the algorithm, the rough mortality rates are smoothed. Using this method, noise is filtered out, which results in a less fluctuations in the data.

The first step in the algorithm is to transform the mortality rates. This is done using the Gompertz transformation. The use of this transformation is derived from Gompertz law of mortality. This law states that the mortality rates of humans increases exponentially with age (see [4]).

The next step is to smooth the transformed mortality rates using kernel smoothing. This will result in less noise in the transformed data. Lastly, using the inversed transformation the smoothed mortality rates are found, which are used in the model.

A.1 Characterisation of the algorithm

The following formulas describe the process of the algorithm in a mathematical sense. The rough mortality rates $Q_{r,x,s}(t^*)$, which are transformed by the following transformation:

$$\hat{f}(x) = \ln(-\ln(1 - Q_{r,x,s}(t^*)))$$

The transformed rates are smoothed using:

$$\min_{a,b,c} \sum_{k=-m}^{m} [f(x + k) - \hat{f}(x + k)]^2$$

where $m = 5$ and $f(x) = a + bx + cx^2$. This means that a second order polynomial is used in the smoothing method.

The resulting $f(x)$ gives the smoothed mortality rates using the transformation:

$$q(x) = 1 - \exp(-\exp(f(x)))$$

The theory, on which every step is based will be explained in the following sections.

A.2 Gompertz transformation

Gompertz law of mortality has proven to accurately describe the mortality rates between ages 30 and 80 in countries where the mortality caused by external factors is low. According to the law, the increase in mortality rates is exponential and can be described using the following cumulative distribution function

$$G(x) = 1 - \alpha e^{-\beta e^{-\gamma x}} - 1$$

where $\alpha, \beta, \gamma \in \mathbb{R}_{>0}$. 

51
The transformation that is used in the algorithm is as follows:

$$\ln(-\ln(1 - Q_r(x)))$$

Observation shows that this will result in a linearization of the data, provided that the data follows the Gompertz function.

![Figure A.1: Gompertz transformed mortality rates for men (black) and women (grey) in 2012. The Gompertz transformation gives an insight in the growth of the mortality rates as age increases.](image)

Figure A.1 shows that the transformed mortality rates are not perfectly Gompertz distributed. However, after age 30 this linear trend is definitely shown. The figure also gives a good insight on how the mortality rates grow along with age. The benefit of the transformation is that the mortality rates do not increase exponentially anymore. This will result in easier smoothing, since the values are more structured.

An extension on the law is the Gompertz-law is the Gompertz-Makeham law, which depends on a age dependent (Gompertz)-factor as well as an age independent factor.

### A.2.1 Gompertz-Makeham law of mortality

The Gompertz-Makeham law is an extension on the Gompertz distribution. The cumulative distribution function of a Gompertz-Makeham distributed population is defined as follows:

$$M(x) = 1 - \exp\left(-\lambda x - \frac{\alpha}{\beta} (e^{\beta x} - 1)\right)$$

for $\alpha, \beta, \lambda \in \mathbb{R}_{>0}$.

Gompertz-Makeham uses an extra parameter $\lambda$. $\lambda$ can be seen as an age indepedent parameter. In general, the Gompertz-Makeham law of mortality gives a better estimation of the
mortality rates than using just Gompertz. However, in the VB algorithm it is not possible to use the Gompertz-Makeham law. The addition of the extra parameter will cause the data to shift when using the transformations. Since it does give a better estimation in general, it would be interesting to find a kernel which could reproduce a Makeham distributed population.

A.3 Kernel smoothing

The algorithm uses kernel smoothing, a non parametric method that estimates a function $f$ using the noisy observations. Since the data does not originate from a parametrized family, a non parametric method must be used to smooth the data. Kernel estimation performs best when the data can be described using a polynomial of order $p < 3$. Fortunately, this has already been achieved by transforming the data.

The basic idea of the kernel smoothing is as follows: The actual value of a datapoint $(x, f(x))$ where $0 \leq x \leq n$, is determined by using a number of neighbouring observations. The observations that are used are also weighted according to their distance to $(x, f(x))$. Since the data points are ages, natural numbers can be used.

In the VB algorithm, the assumption is made that $(x, f(x))$ is part of a underlying polynomial function $f$ of order 2. Using the weighted observations an estimation of $f$ is found. To determine $f$ in the VB algorithm, first the kernel weights must be found. It is also interesting to look at the implicit kernel type that is used.

A.3.1 Weights

To find the actual value of a datapoint $(x, f(x))$, the kernel uses a number of surrounding values. Each of these values is given a certain weight. The assumption is that the transformed mortality rates can be described using a polynomial function of order 2, namely $f(x) = a + bx + cx^2$. The weights can now be determined by looking at the minimalisation function: (also see [2])

$$\min_{a,b,c} \sum_{k=-m}^{m} [f(x+k) - \hat{f}(x+k)]^2$$

Since the minimalisation is only dependent on terms that contain $a, b, c$, the equation can be rewritten as follows:

$$\min \sum_{k=-m}^{m} [f(x+k) - \hat{f}(x+k)]^2$$

$$= \min \sum_{k=-m}^{m} [f(x+k)^2 - 2f(x+k)\hat{f}(x+k)]$$

Substitution of the polynomial and differentiating with respect to $a, b, c$ gives three equations with three unknowns.

Substitution of $f(x) = a + bx + cx^2$ gives:

$$\min \sum_{k=-m}^{m} [f(x+k)^2 - 2f(x+k)\hat{f}(x+k)]$$

$$= \min \sum_{k=-m}^{m} [(a + b(x+k) + c(x+k)^2)^2 - 2(a + b(x+k) + c(x+k)^2)\hat{f}(x+k)]$$

53
Differentiating with respect to \(a, b, c\) results in the following equations:

\[
\sum_{k=-m}^{m} 2(a + b(x + k) + c(x + k)^2) - 2 \sum_{k=-m}^{m} \hat{f}(x + k) = 0
\]

\[
\sum_{k=-m}^{m} 2(a + b(x + k) + c(x + k)^2)(x + k) - 2 \sum_{k=-m}^{m} \hat{f}(x + k)(x + k) = 0
\]

\[
\sum_{k=-m}^{m} 2(a + b(x + k) + c(x + k)^2)(x + k)^2 - 2 \sum_{k=-m}^{m} \hat{f}(x + k)(x + k)^2 = 0
\]

By solving this system of equations, \(a, b, c\) are found. Inserting these into \(f(x) = a + bx + cx^2\) results in the following expression for \(f\):

\[
f(x, h) = \sum_{k=-m}^{m} \left( \frac{3\left(\frac{3}{4}h^2 + \frac{3}{2}h - 1\right) - 15k^2}{(h + 3)(h + 1)(h - 1)} \right) \hat{f}(x + k)
\]

where \(h = 2m\) is chosen as the bandwidth.

Observe that the following properties hold:

\[
\sum_{k=-m}^{m} \left( \frac{3\left(\frac{3}{4}h^2 + \frac{3}{2}h - 1\right) - 15k^2}{(h + 3)(h + 1)(h - 1)} \right) = 1
\]

\[
\sum_{k=-m}^{m} \left( \frac{3\left(\frac{3}{4}h^2 + \frac{3}{2}h - 1\right) - 15k^2}{(h + 3)(h + 1)(h - 1)} \right) k^2 = 0
\]

According to Gavin, J B, Haberman, S and Verrall, R J (11) kernels for which these properties hold, are called optimal smoothing kernels. As a result of these properties, the asymptotic variance of the smoothing kernel is minimized. This means that this specific kernel is an effective one.

Instead of \(h = 2m\), AG uses \(h = 2m + 1\) as a bandwidth, when inserted into equation (A.1) this results in the following formula.

\[
f(qG_{x,s}) = \sum_{k=-m}^{m} \left( \frac{12m^2 + 24m + 5 - 20k^2}{32m(m+1)(m+2)} \right) qG_{x+k,s}
\]

So the weights that are used in the VB algorithm are given by the expression:

\[
\sum_{k=-m}^{m} \frac{12m^2 + 24m + 5 - 20k^2}{32m(m+1)(m+2)}
\]

The algorithm uses different values \(m\) at the early ages. This technique is used because there are not enough data points to use at these ages. The same problem exists for high ages. This, however does not affect the weights.
A.3.2 Bandwidth

The weights that are used in the kernel smoothing are known. Now, a bandwidth must be chosen. The bandwidth determines the number of surrounding values, that are used to determine the value of the datapoint $x_0$. The choice of the kernel bandwidth is therefore very important. A small bandwidth results in a rough estimation of the real data, since OLS is used with few datapoints. A large bandwidth results in a smooth estimation, however this estimation might be inaccurate.

There are some theories about the optimal choice for bandwidth. A commonly used method is using Spencer’s 21-term formula. This formula one of the most succesful ‘moving weighted average’ formulas according to Gavin, Haberman and Verrall (12). For the use of kernels, they state that a bandwidth of 10 approximates Spencer’s formula best and therefore should be used.

This supports the choice of AG, by choosing $m = 5$, which results in a bandwidth of $h = 2m + 1 = 11$. Since one of the terms is the datapoint itself, this results in the use of 10 surrounding datapoints for the estimation.

It also supports the choice of PLT when smoothing the reduction factors. Again, 11 terms are used, which results in the use of 10 surrounding data points.

A.3.3 Kernel Type

Different kernel types exist to smooth data. For example: The Gaussian, Indicator and neighbour kernels. To find out what kernel type is used in the VB algorithm the following equation is rewritten.

$$\min_{a,b,c} \sum_{k=-m}^{m} [f(x + k) - \hat{f}(x + k)]^2$$

$$= \min_{a,b,c} \sum_{k=x-m}^{x+m} [f(k) - \hat{f}(k)]^2$$

$$= \min_{a,b,c} \sum_{k=1}^{n} [f(x - k) - \hat{f}(x - k)]^2 I_{[-m,m]}(k-x)$$

Where $I_{[-m,m]}(k-x)$ is the indicator function, which is 1 when $k - x \in [-m,m]$ and zero otherwise.

Multiplication of the function with the constant $\frac{1}{2m}$ does not change the result of the minimalisation. Besides that, the parameter of the indicator function is also multiplied with this factor, such that a fixed interval is found for the indicator function. This results in:

$$\frac{1}{2m} \min_{a,b,c} \sum_{k=-m}^{m} [f(x + k) - \hat{f}(x + k)]^2$$

$$= \min_{a,b,c} \sum_{k=1}^{n} [f(x) - \hat{f}(x)]^2 \frac{1}{2m} I_{[-\frac{1}{2},\frac{1}{2}]}(\frac{k-x}{2m})$$
A kernel is now found:

$$K_{2m}(x) = \frac{1}{2m} K\left(\frac{x}{2m}\right)$$

Where $K(x) = I_{\left[-\frac{1}{2}, \frac{1}{2}\right]}(x)$ is the so-called indicator kernel. This means that the van Broekhoven algorithm implicitly is based on the indicator kernel to determine weights.
B — Mortality rates for high ages

There is not much useable data for the mortality rates in high ages (100+). This data is not available because there are not many people who survive until these high ages. The oldest person ever in the Netherlands reached an age of 116. When there are so few people the mortality rates are very inaccurate. For example, if there are 5 people of age 108 alive and 4 of them do not reach 109. If the last person reaches the age of 110, we will have \( q_{109} = 0 < q_{108} = 0.8 \). Naturally we expect that \( q_{109} \geq q_{108} \). So we can not properly use this data. Therefore models are constructed to simulate the mortality rates at high ages.

B.1 Gompertz at high ages

Gompertz law states that the mortality rates increase exponentially with age and slowly grows to the asymptotic 1 as a mortality rate. As figure A.1 indicates, the mortality rates after age 35 behave like a Gompertz distribution. Therefore, using Gompertz to model the mortality rates at ages (100+) is a possibility.

However there are also studies that describe a phenomenon called late-life mortality deceleration. This means that at old ages the growth of the mortality rate decreases faster than a Gompertz distribution. These studies therefore state that using the Gompertz function for age (100+) is a poor method. Some specifically high-age models for the mortality rate after age 100 exist, which will be described below.

B.2 Late-life mortality deceleration

When a person grows older, his mortality rate increases (with the exception of the ages 0, . . . , 5). This growth in mortality rates increases rapidly starting around age 35. . . . Lately, there has been a lot of discussion whether or not the growth of mortality rates decreases at high ages. This phenomenon is called late-life mortality deceleration. If this phenomenon holds, it means that the growth of the mortality rates decreases and that they slowly increase to an asymptotic lower than 1.

Studies in the past have stated that the use of the Gompertz function will overestimate the high age mortality rates, since it does not take late-life mortality deceleration into account. They also state that the one-year mortality rates will grow to the asymptotic values 0.439 for women and 0.544.

More recently however, Gavrilov & Gavrilova (2011) stated that the observation of deceleration is caused by distorting factors in the predictions. Examples of these factors are: low data quality or the use of wrong methodologies. Once these distorting factors were dealt with, Gompertz law proved a good fit.
B.3 Kannistö

A model that does take long-life mortality deceleration into account is Kannistö’s model (see: [7]). Using kannistö, the high ages mortality rates can be modelled using a logistic curve, which is also a sigmoid function. This results in a faster decline during the high ages than using Gompertz curve. The Kannistö transformation is as follows:

First the rates are transformed using:

$$\mu_{x,s}(t^*) = -\log(1 - Q_{x,s}(t^*))$$

Then $\mu_{x,s}$ is transformed using logit. This results in:

$$\text{logit}(\mu_{x,s}(t^*)) = \text{logit}(\frac{\mu_{x,s}}{1 - \mu_{x,s}})$$

This results in the following transformed mortality rates.

![Figure B.1: Kannisto transformed mortality rates for men (black) and women (grey) in 2012. The Kannistö transformation gives an insight in the growth of the mortality rates as age increases.](image)

At first, the Figure [B.1] might look identical to Figure [A.1]. The Kannistö transformation again shows a linear trend in the data after age 35. The difference however is in the ages 80-100. The modelling of the mortality rates for the high ages can be done by fitting a linear trend to the transformed data. However, when comparing figure [A.1] with [B.1] a linear trend would fit both models. Therefore, ages 60-100 are examined more closely.
Figure B.2 compares both models. Observation shows that if a linear trend is fitted to the data, Kannistö’s model proves to be more accurate. The values for Gompertz are decreasing a little bit when age 90 is reached. Therefore, consideration of the existence of late life mortality deceleration is justified. This means that the choice of using Kannistö’s model by AG is valid. It is however difficult to check the goodness of Kannistö’s model since there is not enough data to validate this.
C — Data

Many data is used in this project. Information about this data and its source is shown below.

C.1 Proportion of death vs. living

CBS statline: Sterfte; geslacht, leeftijd (op 31 december) en burgerlijke staat.

The data shows the observed amount of deaths and living men and women per age in every year. Using this proportions, the mortality rates and the uncertainty in these rates are constructed.

C.2 Mortality rates

CBS statline: Levensverwachting; geslacht en leeftijd, vanaf 1950 (per jaar).

The data shows the mortality rate at every age in every year, starting in 1950. These mortality rates differ slightly to the proportions calculated with the data used above. This is probably caused due to some corrections CBS could have applied. In this report these are the mortality rates that are used, mainly because of the fact that they are more accurate.

The dataset calculates with a fictual population of 100,000 man and 100,000 women. These fictual men and women are exposed to the given mortality rates, which results in the expected amount of deaths per year per age.

C.3 Average employees age

CBS: Werkzame beroepsbevolking; vergrijzing per bedrijfstak SBI 2008

This data shows the average age of a male and female employee in 2013. The average ages are respectively 42.4 and 40.7 for men and women.

C.4 Interest rates

De Nederlandsche Bank: Nominale rentetermijnstructuur.

The interest rates of DNB shows the risk-free rates for different amount of years investments in every month of the period 1999-2013.

C.5 Insurer fees

Mandema & Partners

The fees of the insurers Delta Lloyd, Nationale Nederlanden and AEGON that are currently used, are provided by Mandema & Partners. Mandema & Partners is an insurance advisory company.
C.6 Maximum age

Gerontology group: Centenarians and supercentenarians

The Gerontology group keeps track of the centenarians and supercentenarians, which are people that live past their 100th and 110th birthday. Using the information about the Netherlands, the maximum age can be determined.
Appendix: Diagonal table

The following figure gives a better insight on how the diagonal table is constructed.

In the fixed table approach the mortality rates in 2035 (light grey) are used to determine $42|p_{25}$. The diagonal table uses the mortality rates in different years (grey) to determine $42|p_{25}$. Obviously, these calculations give different results. So to determine the fee for a 25-year old using the diagonal table, $42|p_{25}$, $43|p_{25}$, $44|p_{25}$, ..., $95|p_{25}$ are determined using a diagonal through the table in figure D.1. The fee for a 26-year old will be calculated the same way, but with a different diagonal (dark grey). Please note, that figure D.1 only shows a part of the diagonal calculation. This can be done for every age, which results in the diagonal table.
Figure D.1: Construction of the diagonal table
E — Appendix: Rcode

AGmodel

1 # CBS statistiek geladen met q_m sterftekans man en q_v sterftekans vrouw. Levensverwachting man, en
2 #ev. levensverwachting vrouw
3 CBSr <- read.delim("C:/Users/Joost/Desktop/BEP/CBSr.txt", dec=";")
4 attach(CBSr)
5
6 source ('C:/Users/Joost/Desktop/BEP/Functions/Qsmoothing.R')
7 source ('C:/Users/Joost/Desktop/BEP/Functions/highagesX.R')
8 source ('C:/Users/Joost/Desktop/BEP/Functions/starttafel.R')
9 source ('C:/Users/Joost/Desktop/BEP/Functions/goaltafel.R')
10 source ('C:/Users/Joost/Desktop/BEP/Functions/Goalsmoothing.R')
11 source ('C:/Users/Joost/Desktop/BEP/Functions/prognose.R')
12 source ('C:/Users/Joost/Desktop/BEP/Functions/levensverwachting.R')
13 source ('C:/Users/Joost/Desktop/BEP/Functions/tarieven.R')
14
15 # AG model
16 # bepalen van de 2- jarige gemiddelde sterftekansen
17
18 QtempM = matrix (, nrow = 126 , ncol = 62)
19 QtempV = matrix (, nrow = 126 , ncol = 62)
20 for (j in 1:62)
21 { for (i in 1:100)
22 { # Qtemp heeft periode in kolommen en leeftijd in rijen
23 QtempM[i,j] = 0.5*(qM[Periode == 1949+j][i]+qM[Periode == 1949+j+1][i])
24 QtempV[i,j] = 0.5*(qV[Periode == 1949+j][i]+qV[Periode == 1949+j+1][i])
25 }
26 }
27
28 #q’s worden de uiteindelijke kansen
29 q_m = matrix (, nrow = 126 , ncol = 62)
30 q_v = matrix (, nrow = 126 , ncol = 62)
31 #smoothing
32 m=4
33 q_m = smoothing(m,QtempM)
34 q_v = smoothing(m,QtempV)
35 #high ages transformation
36 q_m = highages(q_m)
37 q_v = highages(q_v)
38
39 # High ages analysis

65
period = 62
t=1:100
### Gompertz
plot(t-1,log(-log(1-QtempM[t,period])),pch=16,xlab="age",ylab="transformed mortality rates",ylim=c(-10.5,-0.5))
points(t-1,log(-log(1-QtempV[t,period])),pch=16,xlab="age",ylab="transformed mortality rates",col="grey")
### Kannisto
muM = -log(1-QtempM[t,period])
muV = -log(1-QtempV[t,period])
plot(t-1,log(muM/(1-muM)),pch=16,xlab="age",ylab="transformed mortality rates",ylim=c(-10.5,0))
points(t-1,log(muV/(1-muV)),pch=16,xlab="age",ylab="transformed mortality rates",col="grey")
## Kannisto / Gompertz vgl
t=60:100
mu = -log(1-QtempM[1:100,62])
plot(t,log(mu[t]/(1-mu[t])),pch=2,lwd=2,xlab="age",ylab="transformed mortality rates")
points(t,log(mu[t]),pch=16)
# start en goaltafels
# voor kortetermijntrend
jarenkort = 9
qstartM = starttafel(q_m,jarenkort)
qstartV = starttafel(q_v,jarenkort)
# langetermijntrend
jarenlang = 23
thorizon = 50
qgoalM = goaltafel(q_m,qstartM,thorizon,jarenlang)
qgoalV = goaltafel(q_v,qstartV,thorizon,jarenlang)
# smoothen goaltafel + restrictie
l=5
qgoalsM = Goalsmoothing(l,qstartM,qgoalM)
qgoalsV = Goalsmoothing(l,qstartV,qgoalV)
# prognosetafel
qprogM = prognose(q_m,qstartM,qgoalsM,thorizon,jarenkort)
qprogV = prognose(q_v,qstartV,qgoalsV,thorizon,jarenkort)
# EINDE BEREKENING Q_M's
# Levensverwachting
e_m = levensverwachting(qtotaal_m)
e_v = levensverwachting(qtotaal_v)
# EINDE MODEL
for (i in 1:15)
{ 
    temp = smoothing(i,QtempM) 
    residuals=qM[Periode ==2012] - temp[1:100 ,62] 
    q[i] = sum(NA.omit(residuals)*NA.omit(residuals)) 
    w[i] = q[i]/sum(qM[Periode ==2012]) 
}

plot(q,ylab="Sum of squared error",pch=16,lwd=2,xlab="m")
lines(q,lwd=2)

x=60:95
plot(x,QtempM[x,62], xlab="age", type="l",ylab="mortality rates",lwd=2)
lines(x,smoothing(2,QtempM)[x,62], lwd=2,lty=2)
lines(x,smoothing(5,QtempM)[x,62], lwd=2,lty=3)
lines(x,smoothing(15,QtempM)[x,62], lwd=2,lty=4)

# # # # # # # # high ages modelling # # # # # # # # # # #
mu_m = -log(1-q_m)
A = matrix(, nrow = 100 , ncol = 62)
for (i in 81 :95)
{
    A[i ,] = log(mu_m[i ,]/(1-mu_m[i ,]) )
}

Total mortality rate plot Na kannisto
plot(q_m[,62],type="l",lwd=2,ylab="mortality rate",xlab="age")
lines(q_v[,62],lwd=2,lty=2)

Tarieven barplot
v=matrix(0 , nrow = 120 , ncol =68)
for (j in 2:68)
{
    p=NULL
    leeftijd = j
    for (i in 1: leeftijd )
    {
        p[i] = 1
    }

    for (i in leeftijd :120)
    {
        p[i] = (1-q_m[i-1,62])*p[i-1]
    }

    ### Verdisconteren dit
    for (i in 68:120)
    {
        v[i,j]=p[i]/((1.03)^(i- leeftijd))
    }
}
sum(v[68:112,25])
sum(v[68:120,25])
x=68:120
barplot(v[x,25],names.arg=c(x-1),axis.lty=1,space=0.25,xlab="age",ylab="value")
PLTmodel

CBSr <- read.delim("C:/Users/Joost/Desktop/BEP/CBSr.txt", dec="",
source('C:/Users/Joost/Desktop/BEP/Functions/Qsmoothing.R')
source('C:/Users/Joost/Desktop/BEP/Functions/highages.R')
source('C:/Users/Joost/Desktop/BEP/Functions/reductiefactor.R')
source('C:/Users/Joost/Desktop/BEP/Functions/levensverwachting.R')
source('C:/Users/Joost/Desktop/BEP/Functions/tarieven.R')
source('C:/Users/Joost/Desktop/BEP/Functions/glad.R')
source('C:/Users/Joost/Desktop/BEP/Functions/prognosePLT.R')

#CBS model
#1960 t/m 2010 wordt gebruikt bij bepaling schatter

# bepalen van de 2-jarige gemiddelde sterftekansen
Q1tempM = matrix(), nrow = 126, ncol = 63)
Q1tempV = matrix(), nrow = 126, ncol = 63)
for (j in 1:63)
{
  for (i in 1:100)
  {
    #Qtemp heeft periode in kolommen en leeftijd in rijen
    Q1tempM[i,j] = qM[Periode == 1949+j][i]
    Q1tempV[i,j] = qV[Periode == 1949+j][i]
  }
}

#q's worden de uiteindelijke kansen
q1_m = matrix(), nrow = 126, ncol = 63)
q1_v = matrix(), nrow = 126, ncol = 63)

# smoothing
m=5
q1_m = smoothing(m, Q1tempM)
q1_v = smoothing(m, Q1tempV)

# N is jaar
startyear=2012
N = startyear - 1950
thor = 100

aM = reductiefactor(q1_m,N, thor)
aV = reductiefactor(q1_v,N, thor)
agM = glad(aM,5)
agV = glad(aV,5)

# prognose
q1prog_m = prognosePLT(q1_m,agM, thor)
q1prog_v = prognosePLT(q1_v,agV, thor)

#Highages taken from AG
q1totaal_m = data.frame(q1_m, q1prog_m)
q1totaal_v = data.frame(q1_v, q1prog_v)

q1totaal_m[96:121,] = qtotaal_m[96:121,1:ncol(q1totaal_m)]
q1totaal_v[96:121,] = qtotaal_v[96:121,1:ncol(q1totaal_v)]
x = 1958:2062

68
plot(x, levensverwachting(na.omit(qtotaal_m))[68, x-1950], ylim = c(79, 92), type = "l", ylab = "life expectancy", xlab = "year", lwd = 2)

lines(x, levensverwachting(na.omit(qtotaal_m))[68, x-1950], lty = 2, lwd = 2)
lines(x, levensverwachting(na.omit(qtotaal_v))[68, x-1950], col = "grey", lty = 2, lwd = 2)
lines(x, levensverwachting(na.omit(qtotaal_v))[68, x-1950], col = "grey", lwd = 2)

(levensverwachting(na.omit(qtotaal_v))[68, 2062-1950] - levensverwachting(na.omit(qtotaal_m))[68, 2062-1950]) / 2 - (levensverwachting(na.omit(qtotaal_v))[68, 2062-1950] - levensverwachting(na.omit(qtotaal_m))[68, 2062-1950]) / 2
Performance analysis

### Performance analysis. Eerst moeten de smoothed mortality rates worden berekend in de
models.

### Verschillende jaren kunnen gekozen worden. Trend AG kan aangepast worden.

```r
# tot endyear
dendyear = 2000
dend = dendyear - 1950
N = dendyear - 1950
thor = 50
Q1_m = q1totaal_m[1:N]
Q1_v = q1totaal_v[1:N]
aM = reductiefactor(Q1_m, N, thor)
aV = reductiefactor(Q1_v, N, thor)
agM = glad(aM, 5)
agV = glad(aV, 5)

# prognose
Q1prog_m = prognosePLT(Q1_m, agM, thor)
Q1prog_v = prognosePLT(Q1_v, agV, thor)
Highages taken from AG
Q1totaal_m = data.frame(Q1_m, Q1prog_m)
Q1totaal_v = data.frame(Q1_v, Q1prog_v)
Q1totaal_m[96:121,] = qtotaal_m[96:121,1:ncol(Q1totaal_m)]
Q1totaal_v[96:121,] = qtotaal_v[96:121,1:ncol(Q1totaal_m)]
E1_m= levensverwachting(na.omit(Q1totaal_m))
E1_v= levensverwachting(na.omit(Q1totaal_v))

change = q_m
changeV = q_v

for (j in 2:end )
{
  change[,j] = log(q_m[,j-1])/log(q_m[,end])
  changeV[,j] = log(q_v[,j-1])/log(q_v[,end])
}
meanchange = colMeans(change)
meanchangeV = colMeans(changeV)

###### fits hierbij passen en dan kijken welke de kleinste afwijking heeft, #######
error=NULL
errorV=NULL
error[i]=0
error[i]=0

######
for (i in 2:end)
{
  y = i:end
  fit2 <- lm(meanchange[i:end] ~ y)
  linear <- summary(fit2)$coef[1] + summary(fit2)$coef[2]*y
  error[i] = mean((linear - meanchange[i:end])^2)
}

```
y = i:end
fit2 <- lm(meanchangeV[i:end] ~ y)
errorV[i] = mean((linear - meanchangeV[i:end])^2)

plot(1952:endyear, meanchange[2:end], pch=2, ylab="average log mortality rate", xlab="year", ylim=c(0.89, 1.04))
points(1952:endyear, meanchangeV[2:end], pch=20)

x1=26:end
x2=15:end
fit1 = lm(meanchange[x1] ~ x1)
fit2 = lm(meanchange[x2] ~ x2)

# lijntje erin is wel duidelijk
y1 = 20:(end+5)
lines(y1+1950, summary(fit1)$coeff[2]*y1+summary(fit1)$coeff[1], lwd=2, lty=2)
y2 = 10:(end+5)
lines(y2+1950, summary(fit2)$coeff[2]*y2+summary(fit2)$coeff[1], lwd=1.5)

plot(end:2,error[2:end], pch=2, ylab="mean squared error", xlab="duration trend")
points(end:2,errorV[2:end], pch=20)

### data tot 1990
Q_m = qtotal_m[,1:end]
Q_v = qtotal_v[,1:end]

thorizon = 50
jarenkort = 19
jarenlang = 19
Qstart_m = starttafel(Q_m, jarenkort)
Qgoal_m = goaltafel(Q_m, Qstart, thorizon, jarenlang)
Qprognose_m = prognose(Q_m, Qstart_m, Qgoal_m, thorizon, jarenkort)
Qstart_v = starttafel(Q_v, jarenkort)
Qgoal_v = goaltafel(Q_v, Qstart, thorizon, jarenlang)
Qprognose_v = prognose(Q_v, Qstart_v, Qgoal_v, thorizon, jarenkort)
Qtotaal_m = data.frame(Q_m, Qprognose_m[1:121,])
Qtotaal_v = data.frame(Q_v, Qprognose_v[1:121,])

E_m = levensverwachting(na.omit(Qtotaal_m))
E_v = levensverwachting(na.omit(Qtotaal_v))

### Mannen
x=1951:2012
plot(x, E_m[68, x=1950], type="l", ylim=c(79, 83.5), lwd=2, ylab="life expectancy", xlab="year")
points(x:endyear, e_m[68,((x:endyear)-1950)], lwd=2)
points(endyear:2012, e_m[68,((endyear:2012)-1950)], lwd=2, pch=2)
lines(x, E1_m[68, x=1950], type="l", ylim=c(71, 80), lwd=2, lty=2)

### Vrouwen
x=1970:2012
plot(x, E_v[68, x=1950], type="l", lwd=2, ylim=c(81.5,86), lty=1, ylab="life expectancy", xlab="year")
points(x:endyear, e_v[68,((x:endyear)-1950)], lwd=2)
points(endyear:2012, e_v[68,((endyear:2012)-1950)], lwd=2, pch=2)
lines(x, E1_v[68, x=1950], type="l", ylim=c(71, 80), lwd=2, lty=2)
Verzekeraarstarieven

# Verzekeraarstarieven
CBSr <- read.delim("C:/Users/Joost/Desktop/BEP/CBSr.txt", dec=" ")
attach(CBSr)
source('C:/Users/Joost/Desktop/BEP/Functions/Qsmoothing.R')
source('C:/Users/Joost/Desktop/BEP/Functions/highages.R')
source('C:/Users/Joost/Desktop/BEP/Functions/highagesX.R')
source('C:/Users/Joost/Desktop/BEP/Functions/starttafel.R')
source('C:/Users/Joost/Desktop/BEP/Functions/goaltafel.R')
source('C:/Users/Joost/Desktop/BEP/Functions/Goalsmoothing.R')
source('C:/Users/Joost/Desktop/BEP/Functions/tarieven.R')
source('C:/Users/Joost/Desktop/BEP/Functions/levensverwachting.R')
source('C:/Users/Joost/Desktop/BEP/Functions/tarieven.R')
source('C:/Users/Joost/Desktop/BEP/Functions/Schuintafel.R')

DeltaLloyd <- read.delim("C:/Users/Joost/Desktop/BEP/DeltaLloyd.txt")
AEGON <- read.delim("C:/Users/Joost/Desktop/BEP/AEGON.txt", dec=" ")
NatNed <- read.delim("C:/Users/Joost/Desktop/BEP/NN.txt", dec=" ")

## Gegevens inladen
# Tarieven van 16 t/m 67 jaar. DL is 2.5% rr, andere zijn 3%
DL = matrix(, nrow =52 , ncol =2)
NN = matrix(, nrow =52 , ncol =2)
AE = matrix( ,nrow =52 , ncol =2)

attach(DeltaLloyd)
detach(DeltaLloyd)

attach(AEGON)
detach(AEGON)

attach(NatNed)
NN[,1] = Man[1:52]
NN[,2] = Vrouw[1:52]
detach(NatNed)

# Ag model

# bepalen van de 2-jarige gemiddelde sterftekansen
QttempM = matrix(, nrow = 126 , ncol = 62)
QttempV = matrix(, nrow = 126 , ncol = 62)
for (j in 1:62) {
  for (i in 1:100) {
    # Qttemp heeft periode in kolommen en leeftijd in rijen
    QttempM[i,j] = 0.5*(qM[Periode == 1949+j][i]+qM[Periode == 1949+j+1][i])
    QttempV[i,j] = 0.5*(qV[Periode == 1949+j][i]+qV[Periode == 1949+j+1][i])
  }
}

# q's worden de uiteindelijke kansen
q_m = matrix(, nrow = 126 , ncol = 62)
q_v = matrix(, nrow = 126 , ncol = 62)

m=4
q_m = smoothing(m, QtempM)
q_v = smoothing(m, QtempV)

#high ages transformation
q_m = highages(q_m)
q_v = highages(q_v)

#start en goaltafels
#voor kortetermijntrend
jarenkort = 9
qstartM = starttafel(q_m, jarenkort)
qstartV = starttafel(q_v, jarenkort)

#langetermijntrend
jarenlang = 23
thorizon = 100
qgoalM = goaltafel(q_m, qstartM, thorizon, jarenlang)
qgoalV = goaltafel(q_v, qstartV, thorizon, jarenlang)

#smoothen goaltafel+restrictie
l=5
qgoalsM = Goalsmoothing(l, qstartM, qgoalM)
qgoalsV = Goalsmoothing(l, qstartV, qgoalV)

#prognosetafel
qprogM = prognose(q_m, qstartM, qgoalsM, thorizon, jarenkort)
qprogV = prognose(q_v, qstartV, qgoalsV, thorizon, jarenkort)

# ############################################ EINDE BEREKENING Q_M’s

# Levensverwachting
qtotaal_m = data.frame(q_m[1:121, ], qstartM[1:121, ], qgoalM[1:121, ])
qtotaal_V = data.frame(q_v[1:121, ], qstartV[1:121, ], qgoalV[1:121, ])
ex = levensverwachting(qtotaal_m)
plot(ex[2,])

# ############################################ EINDE MODEL

# ################################### TARIEVEN ###########################################
pl = 67
rr = 0.03
corr = 1
netto = 1.005
Fees = na.omit(tarieven(qtotaal_m, pl, rr, corr))
lines(1950:2062, Fees[41, ], type = "l", ylab = "Fee", xlab = "age", ylim = c(3, 6.8))
lines(1950:2062, FeesV[41, ], type = "l", col = "blue")

# Different interest
Fees2 = na.omit(tarieven(qtotaal_m, pl, 0.02, corr))
lines(1950:2062, Fees2[41, ], type = "l", ylab = "Fee", xlab = "year", col = "blue")

Fees3 = na.omit(tarieven(qtotaal_m, pl, 0.01, corr))
Fees4 = na.omit(tarieven(qtotaal_m, pl, 0, corr))
Fees67 = na.omit(tarieven(qtotaal_m, 67, 0.03, corr))
Fees69 = na.omit(tarieven(qtotaal_m, 69, 0.03, corr))
Fees65 = na.omit(tarieven(qtotaal_m, 65, 0.03, corr))

Fees6 = na.omit(tarieven(qtotaal_m, pl, 0, corr))
Fees3 = na.omit(tarieven(qtotaal_m, pl, 0, corr))

# Different PL

x = 1:67
plot(Fees67[x, 62], type = "l", ylab = "Fee", xlab = "age", ylim = c(1, 13))
lines(Fees69[x, 62], type = "l", ylab = "Fee", xlab = "year", col = "orange")
lines(Fees65[x, 62], type = "l", ylab = "Fee", xlab = "year", col = "blue")

# TARIEVEN plotten, vergelijking AG

x = 1:52
plot(x+16, NN[,1], lwd = 2, lty = 1, xlab = "age", ylab = "fee", type = "l", ylim = c(3.2, 16.5))
lines(x+16, AE[,1], lwd = 2, lty = 2)
lines(x+16, DL[,1], lwd = 2, lty = 3)

plot(x+16, NN[,2], lwd = 2, lty = 1, xlab = "age", ylab = "fee", type = "l", ylim = c(3.2, 17.5))
lines(x+16, AE[,2], lwd = 2, lty = 2)
lines(x+16, DL[,2], lwd = 2, lty = 3)

plot(x+16, FeeM[x+16, 90], type = "l", lwd = 2, lty = 1, xlab = "age", ylab = "Fee")
lines(x+16, NN[,1], lwd = 2, lty = 1, col = "grey")
lines(x+16, AE[,1], lwd = 2, lty = 2, col = "grey")
lines(x+16, DL[,1], lwd = 2, lty = 3, col = "grey")

lines(x+16, Fee1M[x+16, 90], type = "l", lwd = 3, lty = 2)
plot(x+16, FeeV[x+16, 91], type = "l", lwd = 2, lty = 1, xlab = "age", ylab = "Fee")
lines(x+16, NN[,2], lwd = 2, lty = 1, col = "grey")
lines(x+16, AE[,2], lwd = 2, lty = 2, col = "grey")
lines(x+16, DL[,2], lwd = 2, lty = 3, col = "grey")

lines(x+16, Fee1V[x+16, 91], type = "l", lwd = 3, lty = 2)

# TARIEVEN plotten, vergelijking PLT

plot(x+16, FeeM[x+16, 90], type = "l", lwd = 2, lty = 1, xlab = "age", ylab = "Fee")
lines(x+16, NN[,1], lwd = 2, lty = 2)
lines(x+16, AE[,1], lwd = 2, lty = 3)
plot(x+16, Fee1M[x+16, 90], type = "l", lwd = 3, lty = 2)
lines(x+16, NN[,2], lwd = 2, lty = 2)
lines(x+16, AE[,2], lwd = 2, lty = 3)
# Schuine tarieven

## Schuine tafels creëren

```r
fairM = NULL
gifV = NULL
gif1M = NULL
gifV = NULL

schuine[1] is een 18-jarige

for (i in 18:68)
{
  fairM[i-17] = FeeM[i,(113+18-i)]
gif1M[i-17] = Fee1M[i,(113+18-i)]
gifV[i-17] = FeeV[i,(113+18-i)]
gif1V[i-17] = Fee1V[i,(113+18-i)]
}

Hoe in kaart te brengen?? Misschien alleen AG laten zien en zeggen PLT ongeveer hetzelfde, of andersom.

gem leeftijd is 41,2. Dus table used moet 2012+25=2037

par(mfrow=c(1,2))
plot(17:67, FeeM[18:68,90], lwd=2, type="l", xlab="age", ylab="Fee", ylim=c(3,15))
lines(17:67, fairM, lwd=2, lty=2)

plot(17:67, Fee1M[18:68,90], lwd=2, type="l", xlab="age", ylab="Fee")
lines(17:67, fair1M, lwd=2, lty=2)

## Rechte tafel in 2035, 1950=1 -> 2039 = 90

par(mfrow=c(1,1))

verschil = NULL
verschil1 = NULL

Tafelkeuze = 90

for (lft in 18:66)
{
  verschil[lft-17] = (sum(FeeM[lft,Tafelkeuze]) + sum(FeeV[lft,Tafelkeuze+1]) - sum(FeeM[lft,(113+18-lft)]) - sum(FeeV[lft,(113+18-lft)]))/ (sum(FeeM[lft,Tafelkeuze]) + sum(FeeV[lft,Tafelkeuze+1]))
  verschil1[lft-17] = (sum(Fee1M[lft,Tafelkeuze]) + sum(Fee1V[lft,Tafelkeuze+1]) - sum(Fee1M[lft,(113+18-lft)]) - sum(Fee1V[lft,(113+18-lft)]))/ (sum(Fee1M[lft,Tafelkeuze]) + sum(Fee1V[lft,Tafelkeuze+1]))
}

barplot(verschil, names.arg=c(18:66), axis.lty=1, xlab="age", ylab="procentual difference")
barplot(verschil1, names.arg=c(18:66), axis.lty=1, xlab="age", ylab="procentual difference")

## Echt schuine tafels

schuinM = schuintafel(data.matrix(qtotaal_m))
gif1M = schuintafel(data.matrix(q1totaal_m))
gifV = schuintafel(data.matrix(qtotaal_v))
gif1V = schuintafel(data.matrix(q1totaal_v))

schuine: 19:67 19-1, 48= 30

(1-Fee1M[49,90]/schuinM[30])*100

par(mfrow=c(1,2))

lft = 21:68
plot(lft, FeeM[lft,90], lwd=2, type="l", xlab="age", ylab="Fee")
lines(lft, fairM[lft-18], lty=2, lwd=2)
lines(lft, schuinM[lft, lwd=2, lty=3)
```
plot(lft,FeeV[lft,91], lwd=2, type="l", xlab="age", ylab="Fee")
lines(lft,fairV[lft-18], lty=2, lwd=2)
lines(lft,schuinV, lwd=2, lty=3)

## PLT model

plot(lft,Fee1M[lft,90], type="l", lwd=2, xlab="age", ylab="Fee")
lines(lft,fair1M[lft-18], lty=2, lwd=2)
lines(lft,schuin1M, lwd=2, lty=3)

plot(lft,Fee1V[lft,91], type="l", lwd=2, xlab="age", ylab="Fee")
lines(lft,fair1V[lft-18], lty=2, lwd=2)
lines(lft,schuin1V, lwd=2, lty=3)

# ###################### Example #############################

KostenMan = matrix(, nrow = 5, ncol =52)
KostenVrouw = matrix(, nrow = 5, ncol =52)

opbouw = 1000

KostenMan[1,] = t(NN[,1]) * opbouw
KostenMan[2,] = t(AE[,1]) * opbouw
KostenMan[3,] = t(DL[,1]) * opbouw
KostenMan[4,] = t(FeeM[17:68,90]) * opbouw
KostenMan[5,] = t(Fee1M[17:68,90]) * opbouw

KostenVrouw[1,] = t(NN[,2]) * opbouw
KostenVrouw[2,] = t(AE[,2]) * opbouw
KostenVrouw[3,] = t(DL[,2]) * opbouw
KostenVrouw[4,] = t(FeeV[17:68,91]) * opbouw
KostenVrouw[5,] = t(Fee1V[17:68,91]) * opbouw

#25
KostenMan[,10]
KostenVrouw[,10]

#45
KostenMan[,30]
KostenVrouw[,30]

#65
KostenMan[,50]
KostenVrouw[,50]

### Verhogen PL wat voor effect

par(mfrow=c(1,1))

Fee65_m = 1.005 * tarieven(na. omit(qtotaal_m),65,0.03,0)
Fee67_m = 1.005 * tarieven(na. omit(qtotaal_m),67,0.03,0)

# change in price

plot(na. omit(Fee65_m[,87]), type="l", lwd=2)
lines(na. omit(Fee67_m[,87]), lwd=2, lty=2)

# development of these rates

x = 1950:2012
y = 2013:2062

plot(1950:2062, colMeans(na. omit(Fee65_m[17:68,])), lwd=2, type="n")
lines(x, colMeans(na. omit(Fee65_m[17:68,x=1949]))), lwd=2)
lines(y, colMeans(na. omit(Fee67_m[17:68,y=1949])), lwd=2)
CBSr <- read.delim("C:/Users/Joost/Desktop/BEP/CBSr.txt", dec="","")
maxage <- read.delim("C:/Users/Joost/Desktop/BEP/maxage.txt")
attach(maxage)
attach(CBSr)
source("C:/Users/Joost/Desktop/BEP/Functions/Qsmoothing.R")
source("C:/Users/Joost/Desktop/BEP/Functions/highages.R")
source("C:/Users/Joost/Desktop/BEP/Functions/starttafel.R")
source("C:/Users/Joost/Desktop/BEP/Functions/goaltafel.R")
source("C:/Users/Joost/Desktop/BEP/Functions/prognose.R")
source("C:/Users/Joost/Desktop/BEP/Functions/levensverwachting.R")

QtempM = matrix(, nrow = 126, ncol = 63)
QtempV = matrix(, nrow = 126, ncol = 63)
for (j in 1:63)
{
  for (i in 1:100)
    # Qtemp heeft periode in kolommen en leeftijd in rijen
    QtempM[i,j] = qM[Periode == 1949+j][i]
    QtempV[i,j] = qV[Periode == 1949+j][i]
}

### Dit zijn de algemene plotjes van levensverwachting, deze doe ik hier omdat de Qtemp
niet getransformeerd worden

x=1:100
plot(x,na.omit(QtempV[,63]),type="l", xlab="age", ylab="mortality rate",lty=1,lwd=2)
lines(x,na.omit(QtempV[,41]), lwd=2,lty=2)
lines(x,na.omit(QtempV[,]1),lwd=2,lty=3)

x=1950:2012
plot(x,levensverwachting(na.omit(QtempM))[1,],xlab="year",ylab="Life expectancy",type="l"
  ,ylim=c(71,84),lwd=2)
lines(x,levensverwachting(na.omit(QtempV))[1,], lwd = 2,lty =2)

### Aantal doden per age per period
NumberM = matrix(0, nrow = 120, ncol = 63)
NumberV = matrix(0, nrow = 120, ncol = 63)
for (j in 1:63)
{
  for (i in 1:100)
    NumberM[i,j]= aantalM[Periode == 1949+j][i]
    NumberV[i,j]= aantalV[Periode == 1949+j][i]
}

## Plotjes voor doden
DifM = matrix(,nrow = 100, ncol =3)
DifV = matrix(,nrow = 100, ncol =3)
DifM[1 ,] = 0
DifV[1 ,] = 0
for (i in 2:100)
{

DifM[i,1] = -(NumberM[i,62] - NumberM[i-1,62])
DifM[i,2] = -(NumberM[i,41] - NumberM[i-1,41])
DifM[i,3] = -(NumberM[i,1] - NumberM[i-1,1])

plot(1:100,DifM[,1],type="l",xlab="age",ylab="Number of deaths",lwd=2,lty=2)
lines(1:100,DifM[,2],lwd =2,lty=3)
lines(1:100,DifM[,3],lwd =1.5 ,lty =1)

fit<-lm(MaxAge~Jaar)
x =1959:2012
plot(Jaar,MaxAge,xlab="Year",ylab="Maximum age",pch =20)
lines(x,(fit$coef[1] +fit$coef[2]*x),lwd =2)

### Waarom smoothen we? schokkerig gedrag van de Q’s laten zien
par(mfrow=c(1,2))
x =1:60
plot(x,na. omit(QtempM)[x ,63],type ="l",xlab = "age",ylab = "mortality rate",lwd =2 )
lines(x,na. omit(QtempV)[x ,63],type ="l",xlab = "age",ylab = "mortality rate",lwd=2,lty =2)
x=60:100
plot(x,na. omit(QtempM)[x ,63],type ="l",xlab = "age",ylab = "mortality rate",lwd =2 )
lines(x,na. omit(QtempV)[x ,63],type ="l",xlab = "age",ylab = "mortality rate",lwd=2,lty =2)
x=80:100
plot(x,na. omit(QtempM)[x ,63],type ="l",xlab = "age",ylab = "mortality rate",lwd =2 )
lines(x,na. omit(QtempV)[x ,63],type ="l",xlab = "age",ylab = "mortality rate",lwd=2,lty =2)

## Mean squared errortje bekijken
## laat een exp fit op i:100 en een lineaire op 1 of 2:i en dan de errors bij elkaar
## optellen, dan vinden we een mooie i
expfits = matrix(,nrow =100,ncol=62)
for (period in 1:62)
{
  thetax = NULL
  thetay = NULL
  thetaz = NULL
  tempy = NULL
  tempx = NULL
  tempz = NULL
  opslag = matrix(10 , nrow =70 , ncol =90)
  for (k in 2:70)
  {
    for (i in k:90)
    {
      x = 1:k
      y = k:i
      z = i:100
      #dalend exp
      logfit0 = lm(log(na. omit(QtempM[x,period]))~x)
      thetax[1] = logfit0$coef[1]
      fit0 = exp(thetax[1]+thetax[2]*x)
      tempx[i] = sum((QtempM[x,period] -fit0)^2)
    }
    tempz = NULL
    #linear
    fit = lm(na. omit(QtempM[y,period])~y)
thetay[1] = fit$coeff[1]

fit1 = thetay[1] + thetay[2]*y
tempty[i] = sum((QtempM[y.period] -fit1)^2)

# stijgend exp
logfit = lm(log(na.omit(QtempM[z.period]))~z)

fit2 = exp(thetaz[1]+thetaz[2]*z)
tempz[i] = sum((QtempM[z.period] -fit2)^2)

opslag[k,i] = tempx[i]+ tempty[i]+ tempz[i]
}
}

minima = arrayInd (which.min(opslag), dim(opslag))

# genaamd sp
X = matrix(), nrow = 2, ncol =2
Y = matrix(), nrow = 2, ncol =2

sp1 = minima[1]
sp2 = minima[2]
x = 1:sp1
y = sp1:sp2
z = sp2:100

fitx = lm(log(na.omit(QtempM[x.period]))~x)
parx1 = summary(fitx)$coeff[1]
parx2 = summary(fitx)$coeff[2]
sigmax1 = summary(fitx)$coeff[1,2]
sigmax2 = summary(fitx)$coeff[2,2]
goodfit0 = exp(parx1+parx2*x)

fity = lm(na.omit(QtempM[y.period]))~y)
pary1 = summary(fity)$coeff[1]
pary2 = summary(fity)$coeff[2]
sigmay1 = summary(fity)$coeff[1,2]
sigmay2 = summary(fity)$coeff[2,2]
goodfit1 = pary1 + pary2*y

fitz = lm(log(na.omit(QtempM[z.period]))~z)
parz1 = summary(fitz)$coeff[1]
parz2 = summary(fitz)$coeff[2]
sigmax1 = summary(fitz)$coeff[1,2]
sigmax2 = summary(fitz)$coeff[2,2]
goodfit2 = exp(parz1+parz2*z)
goodfit = c(goodfit0,goodfit1,goodfit2)
goodfit = goodfit[-sp1]
goodfit = goodfit[-sp2]
expfits[,period] = goodfit
}

plot(levensverwachting(expfits)[68,],type="l",col="blue")
lines(levensverwachting(na.omit(QtempM)))[68,])
lines(levensverwachting(na.omit(q_m)))[68,],col="red")

# ##################################### Einde exp fit #######
plot(QtempM[,1:100,period], xlab ="age", ylab ="mortality rate")
lines(x,goodfit0, col="pink")
```R
lines(y,goodfit1,col="red")
lines(z,goodfit2,col="blue")

### We hebben nu exponentieel gesmoothed voor alle periodes t, nu vergelijken met Qsmoothing ###
par(mfrow=c(1,1))
#plaatje 1:50
x = 1:50
plot(x,na.omit(q_m)[x,period],type="l", xlab = "age", ylab ="mortality rate",lty=2,lwd=2)
lines(x,na.omit(QtempM)[x,period])
lines(x,goodfit[x],lty=3,lwd=2)

y=51:100
plot(y,na.omit(q_m)[y,period],type="l", xlab = "age", ylab ="mortality rate",lwd=2,lty=2)
lines(y,na.omit(QtempM)[y,period])
lines(y,goodfit[y],lwd=2,lty=3)

### Vergelijkingen tussen methodes in harde cijfers
#############################################################
iets=NULL
for (period in 1:62)
{
  a=sum((expfits[1:100,period]-QtempM[1:100,period])^2)
  b=sum((na.omit(q_m)[1:100,period]-QtempM[1:100,period])^2)
  iets[period]= b/a
}
iets>1

exp(parx1)
exp(parx2)
pary1
pary2
exp(parx1)
exp(parx2)

#levensverwachting van een exponential fit
plot(levensverwachting(data.matrix(goodfit,goodfit)),type="l")
lines(levensverwachting(na.omit(QtempM))[,62])
```
## Data analysis

```r
### Onzekerheid in de data?
MannenAbs <- read.delim("C:/Users/Joost/Desktop/BEP/DataAnalyse/MannenAbs.txt", header = FALSE)
MannenRel <- read.delim("C:/Users/Joost/Desktop/BEP/DataAnalyse/MannenRel.txt", header = FALSE, dec = ",")
VrouwenAbs <- read.delim("C:/Users/Joost/Desktop/BEP/DataAnalyse/VrouwenAbs.txt", header = FALSE)
VrouwenRel <- read.delim("C:/Users/Joost/Desktop/BEP/DataAnalyse/VrouwenRel.txt", header = FALSE, dec = ",")
source("C:/Users/Joost/Desktop/BEP/Functions/levensverwachting.R")

### Totale bevolking per leeftijd uitrekenen
Tot_man = data.matrix( (MannenAbs/MannenRel)*1000)
Tot_vrouw = data.matrix( (VrouwenAbs/VrouwenRel)*1000)

### Sterftekansen
Qdata_m = data.matrix(MannenAbs/Tot_man)
Qdata_v = data.matrix(VrouwenAbs/Tot_vrouw)
sigma_man = sqrt((Qdata_m*(1-Qdata_m)/Tot_man)
sigma_vrouw = sqrt((Qdata_v*(1-Qdata_v)/Tot_vrouw)
UP_m = Qdata_m + qnorm(0.975)*sigma_man
LOW_m = Qdata_m - qnorm(0.975)*sigma_man
UP_v = Qdata_v + qnorm(0.975)*sigma_vrouw
LOW_v = Qdata_v - qnorm(0.975)*sigma_vrouw
x = 60:100
plot(x, Qdata_m[x,63], type = "l", lwd=2, lty=3, xlab="age", ylab="mortality rate")
lines(x,UP_m[x,63], lwd=2)
lines(x,LOW_m[x,63], lwd=2)

## what happens to life expectancy when these are taken into account
plot(1950:2012, levensverwachting(LOW_m)[68,,] ,type = "l", lty=2, ylim=c(78,87), xlab="year", ylab="levensverwachting")
lines(1950:2012, levensverwachting(UP_m)[68,,] ,lty=2)
lines(1950:2012, levensverwachting(Qdata_m)[68,,] ,lwd=2)
lines(1950:2012, levensverwachting(UP_v)[68,,] ,lty=2, col="grey")
lines(1950:2012, levensverwachting(LOW_v)[68,,] ,lty=2, col="grey")
lines(1950:2012, levensverwachting(Qdata_v)[68,,] ,lwd=2, col="grey")
```

81
# Shows change of ex when varying short and long trend

CBSr <- read.delim("C:/Users/Joost/Desktop/BEP/CBSr.txt", dec="",)
attach(CBSr)

source('C:/Users/Joost/Desktop/BEP/Functions/Qsmoothing.R')
source('C:/Users/Joost/Desktop/BEP/Functions/highages.R')
source('C:/Users/Joost/Desktop/BEP/Functions/starttafel.R')
source('C:/Users/Joost/Desktop/BEP/Functions/goaltafel.R')
source('C:/Users/Joost/Desktop/BEP/Functions/prognose.R')
source('C:/Users/Joost/Desktop/BEP/Functions/levensverwachting.R')

# bepalen van de 2-jarige gemiddelde sterftekansen

QtempM = matrix(, nrow = 126, ncol = 62)
QtempV = matrix(, nrow = 126, ncol = 62)
for (j in 1:62)
{
  for (i in 1:100)
  {
    # Qtemp heeft periode in kolommen en leeftijd in rijen
    QtempM[i,j] = 0.5*(qM[Periode == 1949+j][i]+qM[Periode == 1949+j+1][i])
    QtempV[i,j] = 0.5*(qV[Periode == 1949+j][i]+qV[Periode == 1949+j+1][i])
  }
}

# q's worden de uiteindelijke kansen
q_m = matrix(, nrow = 126, ncol = 62)
q_v = matrix(, nrow = 126, ncol = 62)

# smoothing
m=4
q_m = smoothing(m, QtempM)
q_v = smoothing(m, QtempV)

# high ages transformation
q_m = highages(q_m)
q_v = highages(q_v)

# start en goaltafels
# voor kortetermijntrend

temp = array(, dim=c(10,113,25))
temp2 = array(, dim=c(10,113,25))
for (j in 1:25)
{
  for (i in 1:10)
  {
    jarenkort = i+5
    qstartM = starttafel(q_m, jarenkort)
    qstartV = starttafel(q_v, jarenkort)
    # langetermijntrend
    jarenlang = j+15
    thorizon = 50
    qgoalM = goaltafel(q_m, qstartM, thorizon, jarenlang)
    qgoalV = goaltafel(q_v, qstartV, thorizon, jarenlang)
  }
}

82
# smoothen goaltafel + restrictie

l=5

qgoalsM = Goalsmoothing(l, qstartM, qgoalM)
qgoalsV = Goalsmoothing(l, qstartV, qgoalV)

# prognosetafel

qprogM = prognose(q_m, qstartM, qgoalsM, thorizon, jarenkort)
qprogV = prognose(q_v, qstartV, qgoalsV, thorizon, jarenkort)
data.frame(qgoalsM, qprogM[1:121,])

data.frame(qgoalsV, qprogV[1:121,])

# EINDE BEREKENING Q_M's

# Levensverwachting

temp[i,j] = levensverwachting(qtotaal_m)[68,]
temp2[i,j] = levensverwachting(qtotaal_v)[68,]

# rij 1 betekent hier i = 6 (komt door de na.omit)

year = 2000
x = year:2063
plot(x, na.omit(temp)[7,(year-1950):(2063-1950),9], ylim=c(75,86), xlab = "year", ylab="Life expectancy",type="l")

lines(x, na.omit(temp)[3,(year-1950):(2063-1950),9], col = "green", lty=1)
lines(x, na.omit(temp)[3,(year-1950):(2063-1950),17], col = "green",lty=2)
lines(x, na.omit(temp)[3,(year-1950):(2063-1950),25], col = "green",lty=3)

lines(x, na.omit(temp)[6,(year-1950):(2063-1950),1], col = "blue",lty=1)
lines(x, na.omit(temp)[6,(year-1950):(2063-1950),17], col = "blue",lty=2)
lines(x, na.omit(temp)[6,(year-1950):(2063-1950),25], col = "blue",lty=3)

lines(x, na.omit(temp)[9,(year-1950):(2063-1950),1], col = "orange",lty=1)
lines(x, na.omit(temp)[9,(year-1950):(2063-1950),17], col = "orange",lty=2)
lines(x, na.omit(temp)[9,(year-1950):(2063-1950),25], col = "orange",lty=3)

lines(x, na.omit(temp)[12,(year-1950):(2063-1950),1], col = "purple",lty=1)
lines(x, na.omit(temp)[12,(year-1950):(2063-1950),17], col = "purple",lty=2)
lines(x, na.omit(temp)[12,(year-1950):(2063-1950),25], col = "purple",lty=3)

lines(x, koning[50:113], col = "blue")

# extremes

temp[6,2062-1950,1]
temp[12,2062-1950,26]
Smoothing

```r
smoothing <- function(l, QtempM)

# smoothen van de 2-jarige gemiddelde sterftekansen
{
  Qs_m = matrix(, nrow = nrow(QtempM), ncol = ncol(QtempM))
  # toewijzing voor x=0,1,2
  Qs_m[1,] = QtempM[1,]
  for (i in 2:3)
  {
    Qs_m[i,] = 0.5 *(QtempM[i-1,]+QtempM[i,])
  }

  Qg_m = matrix(, nrow = nrow(QtempM), ncol(QtempM))
  # toewijzing voor x=3,...,94
  for (i in 1:100)
  {
    Qg_m[i,] = log(-log(1-QtempM[i,]))
  }

  Q_ nieuwM = matrix(, nrow = nrow(QtempM), ncol = ncol(QtempM))
  vm = NULL
  for (j in 1: ncol(QtempM))
  {
    for (i in 4:(99-l-1) )
    {
      m= min(i-3, l)
      for (k in 1:(2*m+2) )
      {
        # Dummy vector v
        vm[k]= Qg_m[i-m-1+k,j]*((12*m^2 + 24*m +5 - 20*(k-1-m-0.5)^2 )/((32*m*(m+1)*(m+2))/3))
      }
      Q_ nieuwM[i,j] = sum(vm)
    }
    for (i in (99-l) :95)
    {
      m=99-i
      for (k in 1:(2*m+2) )
      {
        # Dummy vector v
        vm[k]= Qg_m[i-m-1+k,j]*((12*m^2 + 24*m +5 - 20*(k-1-m-0.5)^2 )/((32*m*(m+1)*(m+2))/3))
      }
      Q_ nieuwM[i,j] = sum(vm)
    }
  }
  # transformatie terug
  for (i in 4: 95)
  {
    Qs_m[i,] = 1-exp(-exp(Q_ nieuwM[i,]))
  }
  return(Qs_m)
}
```
High ages

```r
highages <- function(q_m) {

  mu_m = -log(1-q_m)
  A = matrix(nrow = 100, ncol = 62)
  for (i in 81:95) {
    A[i,] = log(mu_m[i,]/(1-mu_m[i,]))
  }

  # parameters for estimation
  theta1 = NULL
  theta2 = NULL

  x = 81:95
  for (j in 1:62) {
    fit = lm(na.omit(A[,j])~ x)
    theta1[j] = exp(fit$coef[1])
    theta2[j] = fit$coef[2]
  }

  # estimation of high-age mortality rates
  for (j in 1:62) {
    for (i in 96:126) {
      q_m[i,j] = 1-exp((-theta1[j]*exp(i*theta2[j]))/(1+theta1[j]*exp(i*theta2[j])))
    }
  }

  return(q_m)
}
```
levensverwachting <- function(q_m)
{
  #zet sterftekansen om in levensverwachting
  Ix=NULL
  ex=matrix(, nrow = nrow(q_m), ncol= ncol(q_m))
  Ix[1] = 10000
  for (j in 1:ncol(q_m))
  {
    for (i in 2:nrow(q_m))
    {
      Ix[i] = (1-q_m[i,j])*Ix[i-1]
    }
    for (i in 1:nrow(q_m))
    {
      ex[i,j] = sum(Ix[i:nrow(q_m)])/Ix[i] + (i-1)
    }
  }
  return(ex)
}

starttafel <- function(q_m, jaren)
{
  qstart = NULL
  #bepaalt de korte termijn trend van 9 jaar
  for (i in 1:126)
  {
    fkort=(q_m[i,ncol(q_m)]/q_m[i,ncol(q_m)-jaren])^(1/jaren)
    qstart[i] = q_m[i,ncol(q_m)]*fkort
  }
  return(qstart)
}
goaltafel <- function(q_m,qstart,thor, jaren)
{
  qgoal =NULL
  for (i in 1:126)
  {
    flang=(q_m[i,ncol(q_m)]/q_m[i,ncol(q_m)-jaren])^(1/jaren)
    qgoal[i] = qstart[i]^(flang*thor)
  }
  return(qgoal)
Smoothing goaltable

```r
goalsmoothing <- function(l, qgoal, qstart){
  # smoothen van de 2-jarige gemiddelde sterftekansen
  Qgoal = NULL
  # qgoal = qgoalM
  # toewijzing voor x=0,1,2
  for (i in 1:3)
  {
    Qgoal[i]= qgoal[i]
  }
  # Gompertz getransformeerde
  Qg= NULL
  Qg = log(-log(1-qgoal))
  # Gewogen gompertz getransformeerde
  Q_nieuwM = NULL
  vm = NULL
  for (i in 4:(126-l))
  {
    m=min(i-2,1)
    for (k in 1:(2*m+1))
    {
      # Dummy vector v
      vm[k]=Qg[i-m-1+k]*((3*m^2 + 3*m -1 -5*(k-m-1)^2)/((8*m^3 +12*m^2 -2*m -3)/3))
    }
    Q_nieuwM[i] = sum(vm)
  }
  sum(vm)
  # transformatie terug
  for (i in 4:121)
  {
    Qgoal[i] = min(1-exp(-exp(Q_nieuwM[i])),qstart[i])
  }
  Qgoal
  # restrictions
  return(Qgoal)
}
```
Forecast AG

```r
prognose <- function(q_m, qstart, qgoal, thorizon, jaren)
{
  qprog = matrix(, nrow = 126, ncol = thorizon)
  for (i in 1:100)
  {
    fkort = (q_m[i,ncol(q_m)]/q_m[i,ncol(q_m)-jaren])^(1/jaren)
    # bepaling ax voor alle leeftijden
    ax = (log(qgoal[i]) - log(qstart[i]) - thorizon*log(fkort))/((thorizon*(thorizon+1))/2)
    for (j in 1:thorizon)
    {
      # restrictions on qprog
      qprog[i,j] = max(qstart[i]*fkort^j*exp((ax*j*(j +1))/2), qgoal[i])
    }
  }
  for (i in 101:126)
  {
    qprog[i,] = qgoal[i]
  }
  return(qprog)
}
```

Forecast PLT

```r
prognosePLT <- function(q,a, thor)
{
  Q = matrix(,nrow =nrow(q),ncol = thor)
  Q[,1] = (q[,ncol(q)])*a[,1]
  for (j in 2:thor)
  {
    Q[,j]= Q[,j-1]*a[,j]
  }
  return(Q)
}
```
Reduction factors

```r
### Functie bepaald reductiefactor voor CBSmodel
reductiefactor <- function(q1,N,thor)
{
  a=matrix(, nrow= nrow(q1),ncol=thor)
  for (j in 1:ncol(a))
  {
    for (i in 1:nrow(a))
    {
      # max 5 is omdat ze een korte termijn trend willen van minimaal 5
      k=max(5,j-N)
      a[i,j] = (q1[i,N]/q1[i,N-k])^((1/k))
    }
  }
  return(a)
}
```

Smoothing factors

```r
# gladstrijken reductiefactoren
glad <- function(a,l)
{
  v=NULL
  a1= matrix(, nrow = nrow(a),ncol=ncol(a))
  a1[1 ,] = a[1 ,]
  for (j in 1: ncol(a))
  {
    for (i in 2: (nrow(na.omit(a))-l))
    {
      m=min(i-1,l)
      for (k in 1:(2*m+1))
      {
        # Dummy vector 
        v[k]=a[i-m-1+k,j]
      }
      a1[i,j]=mean(v)
    }
    for (i in (nrow(na.omit(a))-l +1) :(nrow(na.omit(a)) -1))
    {
      m=nrow(na.omit(a))-i
      for (k in 1:(2*m+1))
      {
        # Dummy vector
        v[k]=a[i-m-1+k,j]
      }
      a1[i,j]=mean(v)
    }
    a1[nrow(na.omit(a)) ,]=a[nrow(na.omit(a)) ,]
  }
  return(a1)
}
```

89
Modelled fees

tarieven <- function(q_m, pl, rr, correctie) {
  Dx = NULL
  Ix = NULL
  Nx = NULL
  ax = NULL
  dummy = NULL
  Tarief = matrix(nrow = nrow(q_m), ncol = ncol(q_m))
  Ix[1] = 100000
  Dx[1] = 100000
  y = pl + 1 + correctie
  for (j in 1:ncol(q_m)) {
    for (i in 2:nrow(q_m)) {
      Ix[i] = (1-q_m[i,j])*Ix[i-1]
    }
    for (i in 2:nrow(q_m)) {
      Dx[i] = Ix[i]/((1+rr)^((i-1))
    }
    for (i in 1:nrow(q_m)) {
      Nx[i] = sum(Dx[i:nrow(q_m)])
    }
    for (i in 1:(nrow(q_m)-1)) {
      ax[i] = 0.5*(Nx[i]+Nx[i+1])/Dx[i]
    }
    for (i in 1:nrow(q_m)) {
      Tarief[i,j] = dummy[i]*ax[y]
    }
  }
  return(Tarief)
}

Diagonal table

```r
schuintafel <- function(q)
### FUnctie pakt alle waarden schuin en rekent daarmee de bijbehorende tarieven uit.
### keurige methode, er wordt echter gebruikt gemaakt van een prognose van 100 jaar, voor
een 20 jarige. Dat is wel een boel.
{
  Schuin = matrix(0, nrow = 101, 48)
  for (i in 21:68)
  {
    Q=q[i:121,63:(164-i+20)]
    Schuin[(i-20):101,i-20] = diag(Q)
  }

  Nieuw = matrix(0, nrow = 101, ncol = 48)
  Nieuw[1,] = 1
  for (j in 1:48)
  {
    for (i in 2:101)
    {
      Nieuw[i,j] = (1 - Schuin[i,j]) * Nieuw[i-1,j]
    }
  }

  Nieuw[48:101,48]
  toet = matrix(0, nrow = 54, ncol = 48)
  for (j in 1:48)
  {
    for (i in 48:101)
    {
      toet[i-47,j] = (1.03^(-i+1+j))*Nieuw[i,j]
    }
  }

  return(colSums(toet))
}
```